



ADVANCES IN LEGUME RESEARCH

EDITED BY: Diego Rubiales, Susana S. Araújo, Maria C. Vaz Patto,
Nicolas Risipail and Oswaldo Valdés-López
PUBLISHED IN: Frontiers in Plant Science



frontiers

Frontiers Copyright Statement

© Copyright 2007-2018 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-632-1

DOI 10.3389/978-2-88945-632-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ADVANCES IN LEGUME RESEARCH

Topic Editors:

Diego Rubiales, Instituto de Agricultura Sostenible, CSIC, Spain

Susana S. Araújo, Universidade Nova de Lisboa (ITQB NOVA), Portugal

Maria C. Vaz Patto, Universidade Nova de Lisboa (ITQB NOVA), Portugal

Nicolas Rispail, Instituto de Agricultura Sostenible, CSIC, Spain

Oswaldo Valdés-López, Universidad Nacional Autónoma de México (UNAM), Mexico



Legume field trials under the Castle of Almodovar del Rio, Spain.

Image by Diego Rubiales

Legumes crops have an extraordinary importance for the agriculture and the environment. In a world urgently requiring more sustainable agriculture, food security and healthier diets the demand for legume crops is on the rise. The International Legume Society (<http://ils.nsseme.com>) organizes a triannual series of conferences with the goal to serve as a forum to discuss interdisciplinary progress on legume research. The Second International Legume Society Conference (ILS2) hosted in October 2016 at Troia, Portugal was the starting point for the Research Topic “Advances in Legume Research” in FiPS, that was also open to spontaneous submissions.

Citation: Rubiales, D., Araújo, S. S., Patto, M. C. V., Rispail, N., Valdés-López, O., eds. (2018). *Advances in Legume Research*. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88945-632-1

Table of Contents

- 08 Editorial: Advances in Legume Research**
Diego Rubiales, Susana S. Araújo, Maria C. Vaz Patto, Nicolas Risipail and Oswaldo Valdés-López
- 13 Beans (*Phaseolus ssp.*) as a Model for Understanding Crop Evolution**
Elena Bitocchi, Domenico Rau, Elisa Bellucci, Monica Rodriguez, Maria L. Murgia, Tania Gioia, Debora Santo, Laura Nanni, Giovanna Attene and Roberto Papa
- 34 Genetic Diversity of Croatian Common Bean Landraces**
Klaudija Carović-Stanko, Zlatko Liber, Monika Vidak, Ana Barešić, Martina Grdiša, Boris Lazarević and Zlatko Šatović
- 42 Reproductive Strategies in Mediterranean Legumes: Trade-Offs Between Phenology, Seed Size and Vigor Within and Between Wild and Domesticated *Lupinus* Species Collected Along Aridity Gradients**
Jens D. Berger, Damber Shrestha and Christiane Ludwig
- 58 Genotypic Variation in a Breeding Population of Yellow Sweet Clover (*Melilotus officinalis*)**
Kai Luo, M. Z. Z. Jahufer, Fan Wu, Hongyan Di, Daiyu Zhang, Xuanchen Meng, Jiyu Zhang and Yanrong Wang
- 68 Multidisciplinary Contributions to Legume Crop History: Proceed With Caution**
Frank M. Dugan
- 71 Potential of Legume–Brassica Intercrops for Forage Production and Green Manure: Encouragements From a Temperate Southeast European Environment**
Ana M. Jeromela, Aleksandar M. Mikić, Svetlana Vujić, Branko Ćupina, Đorđe Krstić, Aleksandra Dimitrijević, Sanja Vasiljević, Vojislav Mihailović, Sandra Cvejić and Dragana Miladinović
- 78 Influence of Harvest Aid Herbicides on Seed Germination, Seedling Vigor and Milling Quality Traits of Red Lentil (*Lens culinaris* L.)**
Maya Subedi, Christian J. Willenborg and Albert Vandenberg
- 93 Narrow-Leafed Lupin (*Lupinus angustifolius*) β 1- and β 6-Conglutin Proteins Exhibit Antifungal Activity, Protecting Plants Against Necrotrophic Pathogen Induced Damage From *Sclerotinia sclerotiorum* and *Phytophthora nicotianae***
Jose C. Jimenez-Lopez, Su Melser, Kathleen DeBoer, Louise F. Thatcher, Lars G. Kamphuis, Rhonda C. Foley and Karam B. Singh
- 108 Non-host Resistance: DNA Damage is Associated With SA Signaling for Induction of PR Genes and Contributes to the Growth Suppression of a Pea Pathogen on Pea Endocarp Tissue**
Lee A. Hadwiger and Kiwamu Tanaka
- 120 Genome-Wide Dissection of the Heat Shock Transcription Factor Family Genes in *Arachis***
Pengfei Wang, Hui Song, Changsheng Li, Pengcheng Li, Aiqin Li, Hongshan Guan, Lei Hou and Xingjun Wang

- 136 ***Overexpression of Nictaba-Like Lectin Genes From Glycine max Confers Tolerance Toward Pseudomonas syringae Infection, Aphid Infestation and Salt Stress in Transgenic Arabidopsis Plants***
Sofie Van Holle, Guy Smagghe and Els J. M. Van Damme
- 153 ***Understanding the Impact of Drought on Foliar and Xylem Invading Bacterial Pathogen Stress in Chickpea***
Ranjita Sinha, Aarti Gupta and Muthappa Senthil-Kumar
- 166 ***Shoot and Root Traits Contribute to Drought Resistance in Recombinant Inbred Lines of MD 23–24 × SEA 5 of Common Bean***
Jose Polania, Idupulapati M. Rao, Cesar Cajiao, Miguel Grajales, Mariela Rivera, Federico Velasquez, Bodo Raatz and Stephen E. Beebe
- 184 ***Identification of Quantitative Trait Loci Controlling Root and Shoot Traits Associated With Drought Tolerance in a Lentil (Lens culinaris Medik.) Recombinant Inbred Line Population***
Omar Idrissi, Sripada M. Udupa, Ellen De Keyser, Rebecca J. McGee, Clarice J. Coyne, Gopesh C. Saha, Fred J. Muehlbauer, Patrick Van Damme and Jan De Riek
- 195 ***Polyamines Confer Salt Tolerance in Mung Bean (Vigna radiata L.) by Reducing Sodium Uptake, Improving Nutrient Homeostasis, Antioxidant Defense, and Methylglyoxal Detoxification Systems***
Kamrun Nahar, Mirza Hasanuzzaman, Anisur Rahman, Md. Mahabub Alam, Jubayer-Al Mahmud, Toshisada Suzuki and Masayuki Fujita
- 209 ***Association Mapping for Fiber-Related Traits and Digestibility in Alfalfa (Medicago sativa)***
Zan Wang, Haiping Qiang, Haiming Zhao, Ruixuan Xu, Zhengli Zhang, Hongwen Gao, Xuemin Wang, Guibo Liu and Yingjun Zhang
- 216 ***Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement***
Karen M. Frick, Lars G. Kamphuis, Kadambot H. M. Siddique, Karam B. Singh and Rhonda C. Foley
- 228 ***Transcriptome Analysis of a New Peanut Seed Coat Mutant for the Physiological Regulatory Mechanism Involved in Seed Coat Cracking and Pigmentation***
Liyun Wan, Bei Li, Manish K. Pandey, Yanshan Wu, Yong Lei, Liying Yan, Xiaofeng Dai, Huifang Jiang, Juncheng Zhang, Guo Wei, Rajeev K. Varshney and Boshou Liao
- 243 ***Evaluation of Exotically-Derived Soybean Breeding Lines for Seed Yield, Germination, Damage, and Composition under Dryland Production in the Midsouthern USA***
Nacer Bellaloui, James R. Smith, Alemu Mengistu, Jeffery D. Ray and Anne M. Gillen
- 263 ***Overexpression of the Starch Phosphorylase-Like Gene (PHO3) in Lotus japonicus has a Profound Effect on the Growth of Plants and Reduction of Transitory Starch Accumulation***
Shanshan Qin, Yuehui Tang, Yaping Chen, Pingzhi Wu, Meiru Li, Guojiang Wu and Huawu Jiang

- 275 *Arbuscular Mycorrhiza Stimulates Biological Nitrogen Fixation in Two Medicago spp. Through Improved Phosphorus Acquisition***
David Püschel, Martina Janoušková, Alena Voříšková, Hana Gryndlerová, Miroslav Vosátka and Jan Jansa
- 287 *Genotypic Differences in Phosphorus Efficiency and the Performance of Physiological Characteristics in Response to Low Phosphorus Stress of Soybean in Southwest of China***
Tao Zhou, Yongli Du, Shoaib Ahmed, Ting Liu, Menglu Ren, Weiguo Liu and Wenyu Yang
- 300 *Finger on the Pulse: Pumping Iron Into Chickpea***
Grace Z. H. Tan, Sudipta S. Das Bhowmik, Thi M. L. Hoang, Mohammad R. Karbaschi, Alexander A. T. Johnson, Brett Williams and Sagadevan G. Mundree
- 317 *Development of Molecular Markers for Iron Metabolism Related Genes in Lentil and Their Expression Analysis Under Excess Iron Stress***
Debjyoti Sen Gupta, Kevin McPhee and Shiv Kumar
- 325 *Characterization of the Heme Pocket Structure and Ligand Binding Kinetics of Non-symbiotic Hemoglobins From the Model Legume Lotus japonicus***
Laura Calvo-Begueria, Bert Cuypers, Sabine Van Doorslaer, Stefania Abbruzzetti, Stefano Bruno, Herald Berghmans, Sylvia Dewilde, Javier Ramos, Cristiano Viappiani and Manuel Becana
- 339 *Determination of Photoperiod-Sensitive Phase in Chickpea (Cicer arietinum L.)***
Ketema Daba, Thomas D. Warkentin, Rosalind Bueckert, Christopher D. Todd and Bunyamin Tar'an
- 349 *Flowering and Growth Responses of Cultivated Lentil and Wild Lens Germplasm Toward the Differences in Red to Far-Red Ratio and Photosynthetically Active Radiation***
Hai Y. Yuan, Shyamali Saha, Albert Vandenberg and Kirstin E. Bett
- 359 *Comprehensive Analysis of the Soybean (Glycine max) GmLAX Auxin Transporter Gene Family***
Chenglin Chai, Yongqin Wang, Babu Valliyodan and Henry T. Nguyen
- 370 *GmAGL1, a MADS-Box Gene From Soybean, is Involved in Floral Organ Identity and Fruit Dehiscence***
Yingjun Chi, Tingting Wang, Guangli Xu, Hui Yang, Xuanrui Zeng, Yixin Shen, Deyue Yu and Fang Huang
- 381 *De novo Transcriptome Profiling of Flowers, Flower Pedicels and Pods of Lupinus luteus (Yellow Lupine) Reveals Complex Expression Changes During Organ Abscission***
Paulina Glazinska, Waldemar Wojciechowski, Milena Kulasek, Wojciech Glinkowski, Katarzyna Marciniak, Natalia Klajn, Jacek Keszy and Jan Kopcewicz
- 410 *Identification of ZOUPI Orthologs in Soybean Potentially Involved in Endosperm Breakdown and Embryogenic Development***
Yaohua Zhang, Xin Li, Suxin Yang and Xianzhong Feng

- 421 A Conserved Potential Development Framework Applies to Shoots of Legume Species With Contrasting Morphogenetic Strategies**
Lucas Faverjon, Abraham J. Escobar-Gutiérrez, Isabelle Litrico and Gaëtan Louarn
- 435 A Combined Comparative Transcriptomic, Metabolomic, and Anatomical Analyses of Two Key Domestication Traits: Pod Dehiscence and Seed Dormancy in Pea (*Pisum sp.*)**
Iveta Hradilová, Oldřich Trněný, Markéta Válková, Monika Cechová, Anna Janská, Lenka Prokešová, Khan Aamir, Nicolas Krezdorn, Björn Rotter, Peter Winter, Rajeev K. Varshney, Aleš Soukup, Petr Bednář, Pavel Hanáček and Petr Smýkal
- 460 Major Contribution of Flowering Time and Vegetative Growth to Plant Production in Common Bean as Deduced From a Comparative Genetic Mapping**
Ana M. González, Fernando J. Yuste-Lisbona, Soledad Saburido, Sandra Bretones, Antonio M. De Ron, Rafael Lozano and Marta Santalla
- 478 Gene Mapping of a Mutant Mungbean (*Vigna radiata L.*) Using New Molecular Markers Suggests a Gene Encoding a YUC4-like Protein Regulates the Chasmogamous Flower Trait**
Jingbin Chen, Prakrit Somta, Xin Chen, Xiaoyan Cui, Xingxing Yuan and Peerasak Srinives
- 488 A Multiple QTL-Seq Strategy Delineates Potential Genomic Loci Governing Flowering Time in Chickpea**
Rishi Srivastava, Hari D. Upadhyaya, Rajendra Kumar, Anurag Daware, Udit Basu, Philanim W. Shimray, Shailesh Tripathi, Chellapilla Bharadwaj, Akhilesh K. Tyagi and Swarup K. Parida
- 501 Fine-Mapping the Branching Habit Trait in Cultivated Peanut by Combining Bulk Segregant Analysis and High-Throughput Sequencing**
Galya Kayam, Yael Brand, Adi Faigenboim-Doron, Abhinandan Patil, Ilan Hedvat and Ran Hovav
- 512 Phenotyping: Using Machine Learning for Improved Pairwise Genotype Classification Based on Root Traits**
Jiangsan Zhao, Gernot Bodner and Boris Rewald
- 529 Identification of QTL and Qualitative Trait Loci for Agronomic Traits Using SNP Markers in the Adzuki Bean**
Yuan Li, Kai Yang, Wei Yang, Liwei Chu, Chunhai Chen, Bo Zhao, Yisong Li, Jianbo Jian, Zhichao Yin, Tianqi Wang and Ping Wan
- 543 A High-Resolution InDel (Insertion–Deletion) Markers-Anchored Consensus Genetic Map Identifies Major QTLs Governing Pod Number and Seed Yield in Chickpea**
Rishi Srivastava, Mohar Singh, Deepak Bajaj and Swarup K. Parida
- 557 Selection of Novel Cowpea Genotypes Derived Through Gamma Irradiation**
Lydia N. Horn, Habteab M. Ghebrehiwot and Hussein A. Shimelis
- 570 EcoTILLING-Based Association Mapping Efficiently Delineates Functionally Relevant Natural Allelic Variants of Candidate Genes Governing Agronomic Traits in Chickpea**
Deepak Bajaj, Rishi Srivastava, Manoj Nath, Shailesh Tripathi, Chellapilla Bharadwaj, Hari D. Upadhyaya, Akhilesh K. Tyagi and Swarup K. Parida

579 *Gene Classification and Mining of Molecular Markers Useful in Red Clover (*Trifolium pratense*) Breeding*

Jan Ištváněk, Jana Dluhošová, Petr Dluhoš, Lenka Pátková, Jan Nedělník and Jana Řepková

595 *Genotyping-by-Sequencing and its Exploitation for Forage and Cool-Season Grain Legume Breeding*

Paolo Annicchiarico, Nelson Nazzicari, Yanling Wei, Luciano Pecetti and Edward C. Brummer



Editorial: Advances in Legume Research

Diego Rubiales^{1*}, Susana S. Araújo², Maria C. Vaz Patto², Nicolas Rispaill¹ and Oswaldo Valdés-López^{3*}

¹ Institute for Sustainable Agriculture, Consejo Superior de Investigaciones Científicas (CSIC), Córdoba, Spain, ² Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa (ITQB NOVA), Oeiras, Portugal, ³ Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México (UNAM), Tlalnepantla, Mexico

Keywords: legumes, pulses, protein crops, breeding, agronomy, domestication

Editorial on the Research Topic

Advances in Legume Research

In a world urgently requiring more sustainable agriculture, food security and healthier diets the demand for legume crops is on the rise. This growth is fostered by the increasing need for plant protein and for sound agricultural practices that are more adaptable and environmentally sensitive. Food, feed, fibers and even fuel are all products that come from legumes—plants that grow with low nitrogen inputs and in harsh environmental conditions.

To heighten the public awareness of the nutritional benefits of legumes as part of sustainable food production aimed toward food security and nutrition, the 68th UN General Assembly declared 2016 the International Year of Pulses. Timely, the International Legume Society (<http://ils.nsseme.com>) organized the Second International Legume Society Conference (ILS2) from 11 to 14th October 2016 in Portugal serving as an unique forum to present and discuss the new research accomplishments on the legume biology, as well as, seek for innovative scientific approaches intended to address these fundamental questions on this important family of plants. The health and environment benefits, as well as, the marketing of legumes were cross-cutting topics throughout the conference. As a result of contributions made to that conference, the Research Topic “Advances in legume research” (<https://www.frontiersin.org/research-topics/4288/advances-in-legume-research>) was designed, being also open to spontaneous submissions.

Articles published in the Research topic will be now briefly presented in this editorial introduction:

EVOLUTION, CONSERVATION, AND DIVERSITY CHARACTERIZATION

Phaseolus spp. represents an interesting model of crop evolution where five closely related species have been domesticated. Bitocchi et al. reviewed the origin of *Phaseolus* genus, the geographical distribution of the wild species, the domestication process, and the wide spread out of the center of origin. Based on this review, at least seven independent domestication events occurred in the *Phaseolus* genus. This information provides the possibility to unravel the genetic basis of the domestication process not only among species of the same genus, but also between gene pools within the same species. In particular for *P. vulgaris*, this resulted from the breaking of the spatial isolation of the Mesoamerican and Andean gene pools, which allowed spontaneous hybridization, and increasing the possibility of novel genotypes and phenotypes. Accordingly, Carović-Stanko et al. performed a genetic diversity analysis on Croatian common bean landraces and showed that about 27% were of Mesoamerican and 68% of Andean origin, while 4% of the accessions were hybrids between both gene pools.

OPEN ACCESS

Edited by:

Christophe Le May,
Agrocampus Ouest, France

Reviewed by:

Eric Von Wettberg,
University of Vermont, United States
Omer Frenkel,
Agricultural Research Organization,
Israel

*Correspondence:

Diego Rubiales
diego.rubiales@ias.csic.es
Oswaldo Valdés-López
oswaldovaldesl@unam.mx

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 02 February 2018

Accepted: 03 April 2018

Published: 19 April 2018

Citation:

Rubiales D, Araújo SS, Vaz Patto MC,
Rispaill N and Valdés-López O (2018)
Editorial: Advances in Legume
Research. *Front. Plant Sci.* 9:501.
doi: 10.3389/fpls.2018.00501

Understanding adaptive responses to stresses is the key to genetic risk mitigation, by ensuring that crops with an appropriate assemblage of adaptive traits are grown in fitting production niches. Harnessing the genetic diversity of crop plants can be a way to achieve this goal. As a proof-of-concept, Berger et al. investigated wild and domesticated Mediterranean annual reproductive strategies comparing *Lupinus* spp. collected along aridity gradients. Strong trade-offs between seed size, early vigor and phenology were observed. Despite the large differences detected for all these traits, natural and human selections have operated in very similar ways in all species.

Yellow sweet clover (*Melilotus officinalis*) is a legume species widely used as ground cover or green manure. Likewise, this legume has also the potential to be used as forage. However, this last application is not possible due to its inherent high coumarin content. To tackle this limitation, Lou et al. studied the genetic variation for herbage yield, key morphological traits, and coumarin content. Selected populations were polycrossed providing a valuable breeding pool for *M. officinalis* cultivar development in China.

ENVIRONMENT AND AGRONOMY

By knowing the origin and geographic spread of a given crop it is possible to obtain clues about its environmental interactions, including relative adaptation to biotic and abiotic stresses. For instance, archaeobotanical studies have revealed that some legumes (e.g., pea, chickpea, and lentil) are considered as ancient crops of the Fertile Crescent and adjacent areas. However, it is not possible to know the history of legumes in other areas in which data are scarce, like the Bronze Age Steppes. In this respect, Dugan provided an opinion article paper suggesting that an accurate understanding of grain legumes in Proto-Indo-European and Indo-European agriculture and language will have positive consequences for our understanding of archeology, linguistics, and crop biogeography.

There is a renewed interest on intercropping due to its positive effects on crop productivity and resistance against different pathogens. In this line, Jeromela et al. reviewed and highlighted the potential of autumn- and spring-sown intercrops of annual legumes with brassicas for ruminant feeding and green manure.

Lentil producers in northern temperate regions usually apply pre-harvest desiccants as harvest aids to accelerate the lentil crop drying process and facilitate harvesting operations. Despite the beneficial effects of the different pre-harvest aids, there is little information about the impact of the whole battery of harvest aids. Subedi et al. filled this gap up by studying the effect of harvest aids alone or tank mixed with glyphosate on seed germination, seedling vigor, milling, and splitting qualities. This investigation not only confirmed the benefits of using pre-harvest aids, but also showed that the application of diquat alone or in combination with glyphosate improves lentil milling recovery yield.

IMPROVING RESPONSES TO BIOTIC STRESSES

Vicilins are seed storage proteins involved in germination processes supplying amino acids for seedling growth and plant development. Likewise, this type of proteins can have some role in plant defense. To provide more evidence about this, Jimenez-Lopez et al. studied the potential role of β -conglutins from narrow leaf lupin (*Lupinus angustifolius*) in protection against necrotrophic fungal pathogens. Transient expression of $\beta 1$ - and $\beta 6$ -conglutin proteins in *Nicotiana benthamiana* leaves demonstrated *in vivo* growth suppression of *Sclerotinia sclerotiorum* and oomycete *Phytophthora nicotianae*.

The non-host resistance is probably more durable than the single dominant resistance genes; this is because there are diverse types of effectors/elicitors and multiple resistance traits involved. The non-host resistance model with chromatin as a receptor offers flexibility to account for many plant-pathogen interactions. To prove this, Hadwiger and Tanaka studied the aspects of legume defense stimulation by salicylic acid (SA), an inducer of non-host resistance in pea tissue. Authors suggest that the SA-induced PR gene activation may be attributed to the host pea genomic DNA damage.

Heat shock transcription factors (Hsfs) are important transcription factors (TFs) in protecting plant cells from damages caused by various stresses. To increase evidence about the relevance of these TFs in legumes, Wang et al. performed a genome-wide analysis of Hsfs in *Arachis duranensis* and *A. ipaensis*. This analysis led to the identification of at least 16 Hsfs genes in these two *Arachis* species. The identified Hsfs genes were divided in three main groups: A, B, and C. A selection pressure analysis revealed that genes bellowing to the group B underwent a positive selection, while genes bellowing to the group A were affected by a purifying selection. Additionally, the presence of fungal elicitor responsive elements in the promoter region of some of these genes suggests their involvement in response to fungi infection.

Lectins that reside in the nucleocytoplasmic compartment are implicated in plant response to biotic and abiotic stresses. The class of Nictaba-like lectins (NLL) groups all proteins with homology to the tobacco (*Nicotiana tabacum*) lectin, known as stress-inducible lectins. Accordingly, Van Holle et al. showed that soybean NLLs are implicated in stress responses. Overexpression of two soybean NLLs homologs in *Arabidopsis* enhanced tolerance to bacterial infection (*Pseudomonas syringae*), insect infestation (*Myzus persicae*) and salinity.

IMPROVING TOLERANCE TO ABIOTIC STRESSES

In field conditions, plants are exposed to multiple types of stresses. In combined stress scenario, drought can positively or negatively affect pathogen infection. To evaluate this scenario, Sinha et al. investigated the effect of drought stress on interaction of chickpea with *Pseudomonas syringae* pv. *phaseolicola* or *Ralstonia solanacearum*, and the net-effect of combined stress

on chlorophyll content and cell death. This study demonstrated that, regardless of the pathogen, drought-stressed chickpea plants showed a significant reduction in the infection levels. Authors propose that both stress interaction and the net effect of combined stress could be mainly influenced by first occurring stress. Likewise, authors suggest that the outcome of the two-stress interaction in plant depends on both timing of stress occurrence and nature of infecting pathogen.

Drought is one of the major constraints limiting plant growth and yield. Modifications in the root architecture allow plants to increase their water extraction capacity and drought resistance. Another adaptation that plants use to cope with drought is an increase in the photosynthates remobilization to ensure seed production. Thus, specific shoot and root traits seem to be the key in improved resistance to drought in different crop plants. To support this, Polania et al. provide evidence indicating that the resistance to drought stress in common bean (*Phaseolus vulgaris*) is positively associated with a deeper and vigorous root system, better shoot growth, and superior mobilization of photosynthates to pod and seed production. Authors propose that pod harvest index, and seed number per area could serve as useful selection criteria for assessing sink strength and for genetic improvement of drought resistance in common bean.

Specific rooting patterns can be associated with drought avoidance mechanisms that can be used in lentil breeding programs. Idrissi et al. identified QTLs associated with root and shoot traits in a RIL of lentil. A QTL related to root-shoot ratio explained the highest phenotypic variance.

The polyamines (PAs) are low-molecular-weight organic cations found in a wide range of organisms, and perform diverse biological functions. For instance, Nahar et al. studied the physiological roles of PAs for their ability to confer salt tolerance in mung bean seedlings (*Vigna radiata*). This study revealed that exogenous PAs supplementation reduced the salt-induced oxidative stress by increasing the contents of glutathione and ascorbate as well as the activities of glyoxalase enzymes. The overall salt tolerance was reflected through improved water and chlorophyll content, as well as, seedling growth.

IMPROVING FORAGE AND SEED QUALITY

Improved digestibility is a main objective in forage breeding. To achieve this goal, different mapping and association approaches have been used to identify potential trait-marker that can be used for the improvement of digestibility in forage crops. For instance, Wang et al. performed an association mapping analysis in alfalfa (*Medicago sativa*) by genotyping an alfalfa panel and phenotyping for five fiber-related traits in four different environments. From this study, eight associations were predicted in two environments, whereas 20 markers were predicted to be associated with multiple traits. The identification of these trait-marker associations will help to breed alfalfa cultivars with high forage quality.

Quinolizidine alkaloids are toxic secondary metabolites found within the genus *Lupinus*. While they offer the plants protection

against insect pests, their accumulation in grains complicates its use for food purposes as high levels confer a bitter taste and may result in acute anticholinergic toxicity. In this line, Frick et al. discuss possibilities for further elucidation and manipulation of the quinolizidine alkaloids pathway in lupin crops by using conventional and cutting-edge technologies.

Seed-coat cracking and undesirable seed coat color highly affect external appearance and commercial value of peanuts (*Arachis hypogaea*). To face this issue, Wan et al. performed a whole-genome transcriptome analysis on a peanut mutant with cracking and brown colored seed coat (*pscb*). This analysis led to the identification of three candidate genes for the trait, which can be used as marker genes for plant breeding.

High germination, nutritional quality, and yield potential under high heat and dryland production conditions are priority seed traits in soybean (*Glycine max*). Bellaloui et al. searched for these traits in exotic germplasm and identified three breeding lines with consistently superior germination. The study also unveiled potential roles of minerals, especially K, Ca, B, Cu, and Mo, in maintaining high seed quality.

Starch phosphorylase (PHO) catalyses the reversible transfer of glucosyl units from glucose-1-phosphate to the non-reducing ends of α -1,4-D-glucan chains with the release of phosphate. Qin et al. identified three PHO isoform (*LjPHO*) genes in the *Lotus japonicus* genome. Overexpression studies suggested that *LjPHO3* may participate in transitory starch metabolism in *L. japonicus* leaves, but its catalytic properties remain to be studied.

IMPROVING PLANT NUTRITION

Legumes establish root symbioses with rhizobia that provide plants with nitrogen (N) through biological N fixation (BNF), as well as with arbuscular mycorrhizal (AM) fungi that mediate improved plant phosphorus (P) uptake. Püschel et al. studied the interplay between BNF and AM symbioses in *Medicago truncatula* and *M. sativa* along a P-fertilization gradient. The AM symbiosis generally improved P uptake by plants and considerably stimulated the efficiency of BNF under low P availability. In contrast, under high P availability the AM symbiosis brought no further benefits to the plants. Results also suggested competition for limited C resource between the two microsymbionts. The use of P-efficient genotypes is a sustainable management strategy for enhancing yield and P use efficiency. Zhou et al. identified genetic variation for P use efficiency in soybean genotypes under field conditions and studied hydroponically P assimilation characteristics and the related mechanisms of P-efficient soybean genotypes.

Iron deficiency is a major problem in many countries raising interest in biofortification of legumes. Tan et al. reviewed the current status of iron biofortification discussing challenges and potential application of transgenic technology.

Multiple genes and TFs are involved in the uptake and translocation of iron in plants from soil. Sen Gupta et al. developed molecular markers for iron metabolism related genes using genome synteny with *M. truncatula*.

UNDERSTANDING PLANT PHYSIOLOGY

Plant hemoglobins (Hbs) are found in nodules of legumes and actinorhizal plants but also in non-symbiotic organs of monocots and dicots. Non-symbiotic Hbs (nsHbs) have been classified into two phylogenetic groups. Class 1 nsHbs show an extremely high O₂ affinity and are induced by hypoxia and nitric oxide (NO), whereas class 2 nsHbs have moderate O₂ affinity and are induced by cold and cytokinins. Using spectroscopic analyses, Calvo-Begueria et al. showed major differences between the two phylogenetic classes of nsHbs and also between the two members of the same class, strongly suggesting that these three globins perform non-redundant functions.

Photoperiod is one of the major environmental factors determining time to flower initiation and first flower appearance in plants. Daba et al. studied photoperiod sensitivity in chickpea (*Cicer arietinum*). Photoperiod-sensitive and -insensitive phases were identified by experiments in which individual plants were reciprocally transferred in a time series from long to short days and vice versa in growth chambers. Results from this research will help to develop cultivars with shorter pre-inductive photoperiod-insensitive and photoperiod-sensitive phases to fit to regions with short growing seasons.

Light is essential for plant growth and development. Yuan et al. studied the response of cultivated lentil and wild *Lens* germplasm to different light environments, showing that days to flower of *Lens* genotypes were mainly influenced by changes in the red/far-red ratio of the light quality but not by changes in the intensity of the photosynthetically active radiation. The distinctly different responses between flowering time and elongation under low red/far-red conditions among wild *Lens* genotypes suggest that flowering and elongation are controlled by discrete pathways. Three *L. lamottei* and one *L. ervoides* genotypes were less sensitive to changes in light quality maintaining similar yield, biomass, and harvest index across all three light environments; these are indications of better adaptability toward changes in light environment.

The phytohormone auxin plays also a critical role in regulation of plant growth and development as well as in plant responses to abiotic stresses. Auxin transporters are major players in polar auxin transport. Chai et al. performed a genome-wide analysis of the soybean *GmLAX* auxin transporter gene family. A total of 15 *GmLAX* genes were identified in the soybean genome distributed on 10 out of the 20 soybean chromosomes. *GmLAX*s showed very dynamic expression patterns, most of them responsive to drought, salt and dehydration stresses, as well as, auxin and abscisic acid stimuli, in a tissue- and/or time sensitive manner.

MADS-domain proteins are important TFs involved in many aspects of plant reproductive development. Chi et al. found that *GmAGL1* is specifically expressed in reproductive tissues but not in roots, stems, and leaves of soybean. The ectopic overexpression of *GmAGL1* in *Arabidopsis* suggested a role for this MADS-box protein in floral organ identity and fruit dehiscence.

Excessive flower and pod abscission represents an economic drawback for yellow lupine (*Lupinus luteus*). Glazinska et al. studied transcriptional networks in the pods, flowers and flower

pedicels to identify genes playing key roles in generative organ abscission in yellow lupine. Auxin, ethylene and gibberellins were some of the main factors engaged in generative organ abscission. Identified differentially expressed genes common for all library comparisons were involved in cell wall functioning, protein metabolism and water homeostasis and stress response.

Improvement of seed quality requires deep insights into the genetic regulation of seed development. The endosperm serves as a temporary source of nutrients that are transported from maternal to filial tissues. It also generates signals for proper embryo formation. Zhang et al. showed that soybean *GmZOU-1* gene is an ortholog of the *Arabidopsis* bHLH domain TF that may be responsible for endosperm breakdown and embryo cuticle formation in soybean.

ADJUSTING PLANT GROWTH AND DEVELOPMENT

Plant morphology markedly affects its competitive ability and persistence in mixtures. Faverjon et al. compared the patterns of shoot organogenesis and shoot organ growth in contrasting forage species belonging to the four morphogenetic groups (i.e., stolon-formers, rhizome-formers, crown-formers tolerant to defoliation and crown-formers intolerant to defoliation). The consequences of this quantitative framework are discussed, along with its possible applications regarding plant phenotyping and modeling.

Seed dispersal and germination are two key traits that have been selected to facilitate cultivation and harvesting of crops. Hradilová et al. studied anatomical structure of seed coat and pod, and identified metabolic compounds associated with water-impermeable seed coat and differentially expressed genes involved in pea (*Pisum* spp.) seed dormancy and pod dehiscence. This integrated analysis of seed coat in wild and cultivated pea provides new insight as well as raises new questions associated with domestication, seed dormancy and pod dehiscence.

Determinacy growth habit and accelerated flowering are adaptive traits in common bean. Through a comparative mapping approach, González et al. detected additive and epistatic QTLs regulating flowering time, vegetative growth, and rate of plant production. Further QTL analysis coupled with previous results highlighted *001G189200* gene, homologous to the *Arabidopsis thaliana* *TFL1* gene, as a candidate gene for determinacy locus.

Mungbean (*Vigna radiata*) is a cleistogamous plant in which flowers are pollinated before they open, which prevents yield improvements through heterosis. Chasmogamous mutant (CM) plants are available. Chen et al. mapped the *cha* gene responsible for the chasmogamous flower trait to a 277.1-kb segment on chromosome 6. Twelve candidate genes were detected in this segment, including *Vradi06g12650*, which encodes a YUCCA family protein associated with floral development. A single base pair deletion producing a frame-shift mutation and a premature stop codon in *Vradi06g12650* was detected only in CM plants suggesting *Vradi06g12650* as *cha* candidate gene.

The time of flowering has a major influence in both plant productivity and adaptation to the changing environment. Hence, different efforts to understand the genetic programs underlying the flowering time in important crop plants have been made. For instance, Srivastava et al. used a high-throughput multiple QTL-seq strategy to identify two major QTL genomic regions governing flowering time on chickpea (*Cicer arietinum*) chromosome 4. The functionally relevant molecular tags delineated can be used for deciphering the natural allelic diversity-based domestication pattern of flowering time and expediting genomics-aided crop improvement to develop early flowering cultivars of chickpea.

The branching habit is an important descriptive and agronomic character of peanut (*Arachis hypogaea*). Kayam et al. fine-mapped this trait by combining high-throughput sequencing and bulk segregant analysis, providing a baseline for candidate gene discovery and map-based cloning.

Proper phenotyping is becoming a bottleneck in breeding, being particularly difficult for inherent root system architectures. To overcome this constrain, Zhao et al. provided machine learning algorithms that were used for unbiased identification of most distinguishing root traits and subsequent pairwise pea (*Pisum sativum*) cultivar differentiation.

Fine mapping of quantitative trait loci (QTL) and qualitative trait genes plays an important role in gene cloning, molecular-marker-assisted selection (MAS), and trait improvement. As a proof-of-concept, Li et al. mapped 26 agronomic QTLs and five qualitative trait genes related to pigmentation in adzuki bean (*Vigna angularis*). For this mapping analysis, authors used 1,571 polymorphic SNP markers generated via Restriction-site-Associated DNA sequencing technology. The identification of these QTL and qualitative trait genes will contribute to breed adzuki bean cultivars with desirable traits.

Another example about the relevance of the QTL analysis in plant breeding programs is provided by Srivastava et al. who used a high-throughput whole genome next-generation resequencing strategy to develop InDel markers in chickpea (*Cicer arietinum*) mapping populations. By using this approach, Srivastava et al. identified three major QTLs governing pod number and seed yield per plant. These functionally relevant molecular tags can drive marker-assisted genetic enhancement to develop high-yielding cultivars with increased seed/pod number and yield in chickpea.

GENOMIC TOOLS FOR BREEDING

Genetic variation is the basis for plant breeding programs. Most conventional crop improvement programs rely on natural genetic variation present among germplasm pools. Alternatively, induced mutagenesis still offers the potential to create valuable genetic variation for genetic enhancement and breeding. For instance, Horn et al. identified agronomically desirable cowpea (*Vigna unguiculata*) mutants after gamma irradiation. Ten phenotypically and agronomically stable novel mutants are

described that constitute a valuable genetic resource for cowpea genetic enhancement and breeding.

The large-scale mining and high-throughput genotyping of novel gene-based allelic variants in natural mapping population are essential for association mapping to identify functionally relevant molecular tags governing useful agronomic traits. For instance, Bajaj et al. used an EcoTILLING approach coupled with agarose gel detection assay to discover 1133 novel SNP allelic variants by genotyping in a *desi* and *kabuli* chickpea collection constituting a seed weight association panel. Integrating genotyping and phenotyping data identified eight SNP alleles in the eight TF genes regulating seed weight of chickpea.

Ištvánek et al. used illumina paired-end sequencing of red clover (*Trifolium pratense*) allowing the identification of large sets of SSRs and SNPs throughout that will be key for implementing genome-based breeding approaches, for identifying genes underlying key traits, and for genome-wide association studies.

Genotyping-by-Sequencing (GBS) may drastically reduce genotyping costs compared with SNP array platforms. Annicchiarico et al. compared GBS protocols on legume species that differ for genome size, ploidy, and breeding system, and showed successful applications and challenges of GBS data on legume species. Authors devised a simple method for comparing phenotypic vs. genomic selection in terms of predicted yield gain per year for same evaluation costs, whose application to preliminary data for alfalfa and pea in a hypothetical selection scenario for each crop indicated a distinct advantage of genomic selection.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

The organizers of the Second International Legume Society Conference (<http://www.itqb.unl.pt/meetings-and-courses/legumes-for-a-sustainable-world>) are acknowledged as the idea of this Research Topic arose from the conference. However, the RT was not restricted to presentations made at the conference but was also open to other relevant quality spontaneous submission.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Rubiales, Araújo, Vaz Patto, Risipail and Valdés-López. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Beans (*Phaseolus* ssp.) as a Model for Understanding Crop Evolution

Elena Bitocchi¹, Domenico Rau², Elisa Bellucci¹, Monica Rodriguez², Maria L. Murgia², Tania Gioia³, Debora Santo¹, Laura Nanni¹, Giovanna Attene² and Roberto Papa^{1*}

¹ Department of Agricultural, Food and Environmental Sciences, Marche Polytechnic University, Ancona, Italy, ² Department of Agriculture, University of Sassari, Sassari, Italy, ³ School of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, Potenza, Italy

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Consejo Superior de Investigaciones
Científicas (CSIC), Spain

Reviewed by:

Zlatko Satovic,
University of Zagreb, Croatia
Klaudija Carović-Stanko,
University of Zagreb, Croatia

*Correspondence:

Roberto Papa
r.papa@univpm.it

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 31 December 2016

Accepted: 19 April 2017

Published: 08 May 2017

Citation:

Bitocchi E, Rau D, Bellucci E,
Rodriguez M, Murgia ML, Gioia T,
Santo D, Nanni L, Attene G and
Papa R (2017) Beans (*Phaseolus*
ssp.) as a Model for Understanding
Crop Evolution.
Front. Plant Sci. 8:722.
doi: 10.3389/fpls.2017.00722

Here, we aim to provide a comprehensive and up-to-date overview of the most significant outcomes in the literature regarding the origin of *Phaseolus* genus, the geographical distribution of the wild species, the domestication process, and the wide spread out of the centers of origin. *Phaseolus* can be considered as a unique model for the study of crop evolution, and in particular, for an understanding of the convergent phenotypic evolution that occurred under domestication. The almost unique situation that characterizes the *Phaseolus* genus is that five of its ~70 species have been domesticated (i.e., *Phaseolus vulgaris*, *P. coccineus*, *P. dumosus*, *P. acutifolius*, and *P. lunatus*), and in addition, for *P. vulgaris* and *P. lunatus*, the wild forms are distributed in both Mesoamerica and South America, where at least two independent and isolated episodes of domestication occurred. Thus, at least seven independent domestication events occurred, which provides the possibility to unravel the genetic basis of the domestication process not only among species of the same genus, but also between gene pools within the same species. Along with this, other interesting features makes *Phaseolus* crops very useful in the study of evolution, including: (i) their recent divergence, and the high level of collinearity and synteny among their genomes; (ii) their different breeding systems and life history traits, from annual and autogamous, to perennial and allogamous; and (iii) their adaptation to different environments, not only in their centers of origin, but also out of the Americas, following their introduction and wide spread through different countries. In particular for *P. vulgaris* this resulted in the breaking of the spatial isolation of the Mesoamerican and Andean gene pools, which allowed spontaneous hybridization, thus increasing of the possibility of novel genotypes and phenotypes. This knowledge that is associated to the genetic resources that have been conserved *ex situ* and *in situ* represents a crucial tool in the hands of researchers, to preserve and evaluate this diversity, and at the same time, to identify the genetic basis of adaptation and to develop new improved varieties to tackle the challenges of climate change, and food security and sustainability.

Keywords: domestication, genetic diversity, adaptation, population genomics, crop evolution, convergent evolution, selection signatures

INTRODUCTION

Beans (*Phaseolus* spp.), and in particular the common bean *P. vulgaris* L., represent the most important grain legume for direct human consumption worldwide. They are a major source of highly valuable plant protein and micronutrients (Broughton et al., 2003; Vaz Patto et al., 2015), they provide health benefits that are related to their regular consumption (Messina, 2014; Bitocchi et al., 2016b), and they contribute to sustainable improvements to the environment when they are grown in agricultural rotation or with intercropping, due to their biological nitrogen fixation, their effects on the soil, and their control of weeds (Rubiales and Mikic, 2015; Bitocchi et al., 2016b). Thus, beans have a key role in the diversification and sustainable intensification of agriculture, particularly in light of the new and urgent challenges, such as climate change. However, it has to be considered that without a deep knowledge of the evolutionary history of crops, no improvements are possible, inasmuch as evolutionary studies provide breeders with information about the available genetic diversity and the genetic control of important agronomic traits related to adaptation and domestication (Bitocchi et al., 2016b).

As Charles Darwin suggested, crop domestication can be seen as a “giant experiment” to test the evolutionary hypothesis. During domestication, similar sets of traits were selected over a wide range of plant species, as the so-called domestication syndrome, which shows numerous examples of convergent phenotypic evolution. *Phaseolus* ($2n = 2X = 22$) is a unique example of multiple parallel and independent domestications. Indeed, not only did domestication occur in five closely related species, *Phaseolus vulgaris*, *P. lunatus*, *P. coccineus*, *P. dumosus* (formerly *P. polyanthus*), *P. acutifolius*, but also, different from other crop species, both *P. vulgaris* and *P. lunatus* have undergone two independent domestications. One was in Mesoamerica and the other in the Andes, which occurred during their reproductive isolation that was caused by the geographical barriers between these gene pools. Thus, considering single independent domestication events for *P. coccineus*, *P. dumosus* and *P. acutifolius*, and two for both *P. vulgaris* and *P. lunatus*, at least seven independent and isolated processes of domestication have occurred for *Phaseolus*.

A different question that arises is the occurrence of multiple domestications within species or within gene pools, where after domestication there was a lack of, or there were incomplete, reproductive barriers, with gene flow occurring among early domesticates, as can also be seen in for other crops (Meyer et al., 2012). In most of these cases where multiple domestications have occurred, this included gene flow between the early domesticates, while for beans, the strong geographical isolation between the gene pools guaranteed their reproductive isolation. However, this does not exclude *per se* the occurrence of multiple domestications within each gene pool of *P. lunatus* and *P. vulgaris*, and this topic is discussed later in this review. The only example that might be similar to *Phaseolus* was the domestication of rice, with the independent domestication of the *indica* and *japonica* subspecies in Asia (Vitte et al., 2004; Londo et al., 2006). However, recent studies appear to demonstrate a more complex situation for

rice, which involved a single domestication with the subsequent divergence of these two subspecies (Molina et al., 2011; Choi et al., 2017).

The *Phaseolus* genus also represents a very interesting model because the divergence among these five *Phaseolus* spp. is relatively recent (2–4 Ma ago; Delgado-Salinas et al., 2006). Indeed, genomic and cytogenetic analyses have reported high levels of collinearity and synteny among the *Phaseolus* genomes (Bonifácio et al., 2012; Fonsêca and Pedrosa-Harand, 2013; Gujaria-Verma et al., 2016), which suggests conserved gene function.

The *Phaseolus* spp. have different breeding systems and life history traits, from annual and autogamous, to perennial and allogamous. This provides the opportunity to determine whether these features have direct consequences on the effects of domestication on the phenotypic and genotypic architecture of the crop plants. Another important aspect, at least for the common bean, is the complex pattern of expansion and the pathways of distribution out of the American domestication centers. This also involved several introductions from the New World that were combined with exchanges between continents, and among several countries within continents. Thus, along with a dramatic increase in the amplitude of the agro-ecological conditions for these crops, new genetic combinations have allowed new opportunities for natural and human-mediated selection that might have promoted their adaptation to specific environmental conditions.

All of these features make *Phaseolus* spp. an ideal model to study domestication and evolution. Thus, the present review offers an overview of the current knowledge of the evolutionary history of the *Phaseolus* crop species, with particular focus on *P. vulgaris* L. and on the recent outcomes relating to the genetic bases of important domestication and adaptation traits.

ORIGIN OF THE SPECIES OF THE *Phaseolus* GENUS

Among the ~70 species that belong to the *Phaseolus* genus, most are geographically distributed in Mesoamerica (Figure 1), where the genus appears to have diversified within the past 4–6 Ma ago (Delgado-Salinas et al., 2006). This diversification of the different species is likely to have taken place during and after the tectonic events that led to the present-day form of Mexico (Delgado-Salinas et al., 2006), which appeared in the Late Miocene (5 Ma ago; Nieto-Samaniego et al., 1999; Alva-Valdivia et al., 2000). In particular, phylogenetic analyses have shown that the *Phaseolus* spp. can be grouped into two major sister clades: clade A, which comprises the Pauciflorus, Pedicellatus, and Tuerckheimii groups, and the weakly resolved species (i.e., *P. glabellus*, *P. macrolepis*, *P. microcarpus*, and *P. oaxacanus*); and clade B, which comprises the Filiformis, Vulgaris, Lunatus, Leptostachyus, and Polystachios groups (Delgado-Salinas et al., 2006). Thus, eight principal crown clades that show some morphological, ecological and bio-geographic distinctions characterize the *Phaseolus* genus, and their formation occurred relatively late on, with an average age of ~2 Ma ago

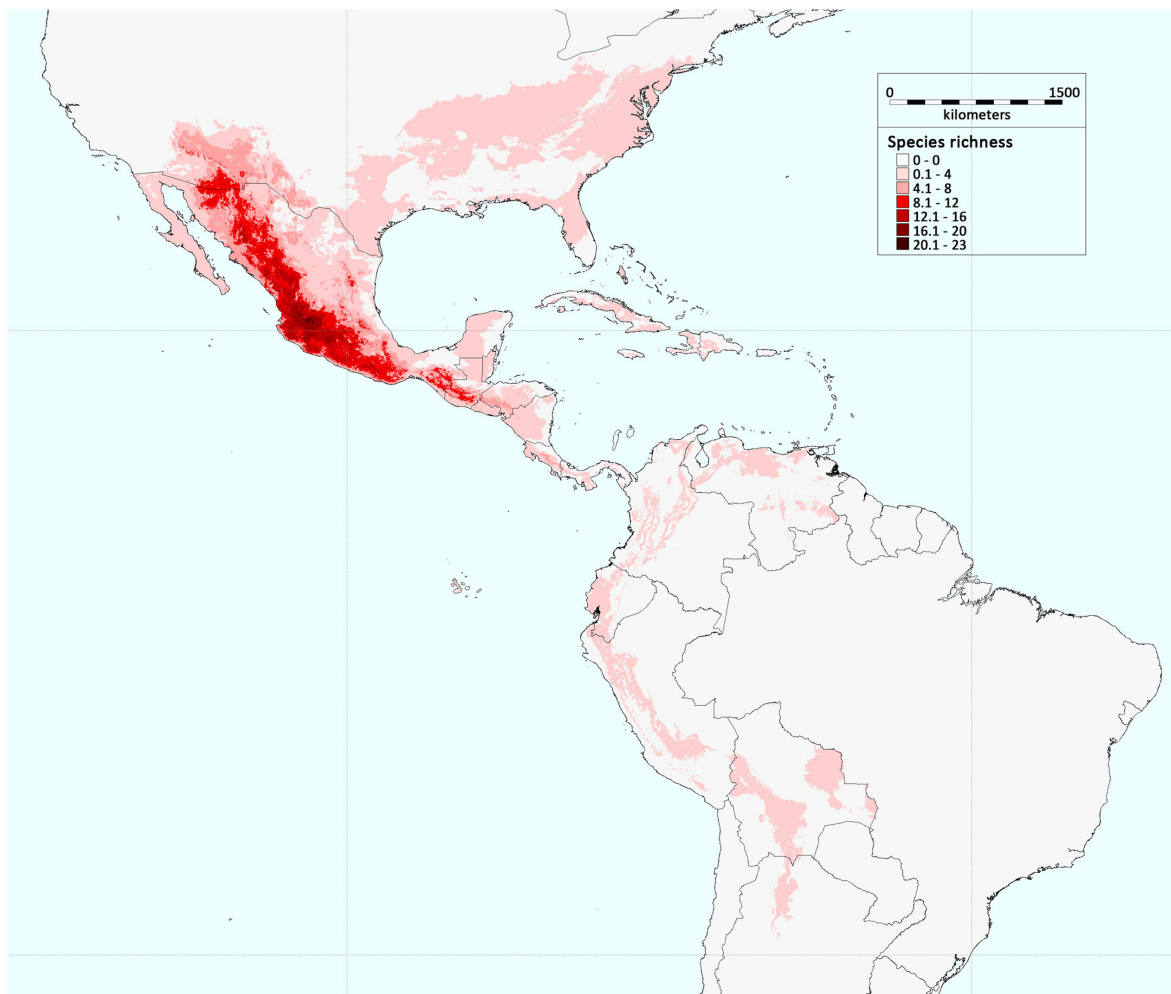


FIGURE 1 | Geographic distribution of *Phaseolus* species richness in America. The colors in the legend indicate the number of *Phaseolus* species present in a defined geographic area (modified from Ramírez-Villegas et al., 2010; permissions for reproduction have been obtained from the copyright holders).

(Delgado-Salinas et al., 2006). The oldest group is *Vulgaris*, which has been dated at ~ 4 Ma ago (Delgado-Salinas et al., 2006).

Five *Phaseolus* species were domesticated: the common bean *P. vulgaris*; the year bean *P. dumosus* Macfad.; the runner bean *P. coccineus* L.; the tepary bean *P. acutifolius* A. Gray; and the Lima bean *P. lunatus* L.. The Lima bean belongs to the *Lunatus* group, the formation of which has been dated at ~ 2 Ma ago (Delgado-Salinas et al., 2006); it is a predominantly autogamous species that includes both annual determinate bush types and indeterminate climbers that are often perennials, due to their enlarged tap root (Baudoin, 1988; Salunkhe and Kadam, 1998) (Figure 2). All the other *Phaseolus* crop species (i.e., *P. vulgaris*, *P. dumosus*, *P. coccineus*, and *P. acutifolius*) belong to the *Vulgaris* group. In particular, *P. vulgaris*, *P. dumosus*, and *P. coccineus* are very closely related, and these three species are partially intercrossable, although only when *P. vulgaris* is the female parent (Mendel, 1866; Wall, 1970; Shii et al., 1982; Hucl and Scoles, 1985). This is despite their marked differences in mating systems and life cycles, as *P. vulgaris* is predominantly autogamous

and annual, *P. coccineus* is predominantly allogamous and perennial, and *P. dumosus* has intermediate characteristics between *P. coccineus* and *P. vulgaris* (Figure 2). Analysis of sequence data of the α -amylase inhibitor gene indicated that *P. vulgaris* diverged from *P. dumosus* and *P. coccineus* ~ 2 Ma ago (Gepts et al., 1999). *P. acutifolius*, is an annual species with a highly selfing reproductive system.

GEOGRAPHIC DISTRIBUTION, ORIGIN AND ADAPTATION OF THE WILD FORMS OF DOMESTICATED *PHASEOLUS* SPECIES

Geographic Distribution and Origin

The wild forms of *P. vulgaris* and *P. lunatus* are distributed in both Mesoamerica and South America, while those of *P. dumosus*, *P. coccineus*, and *P. acutifolius* have a geographic

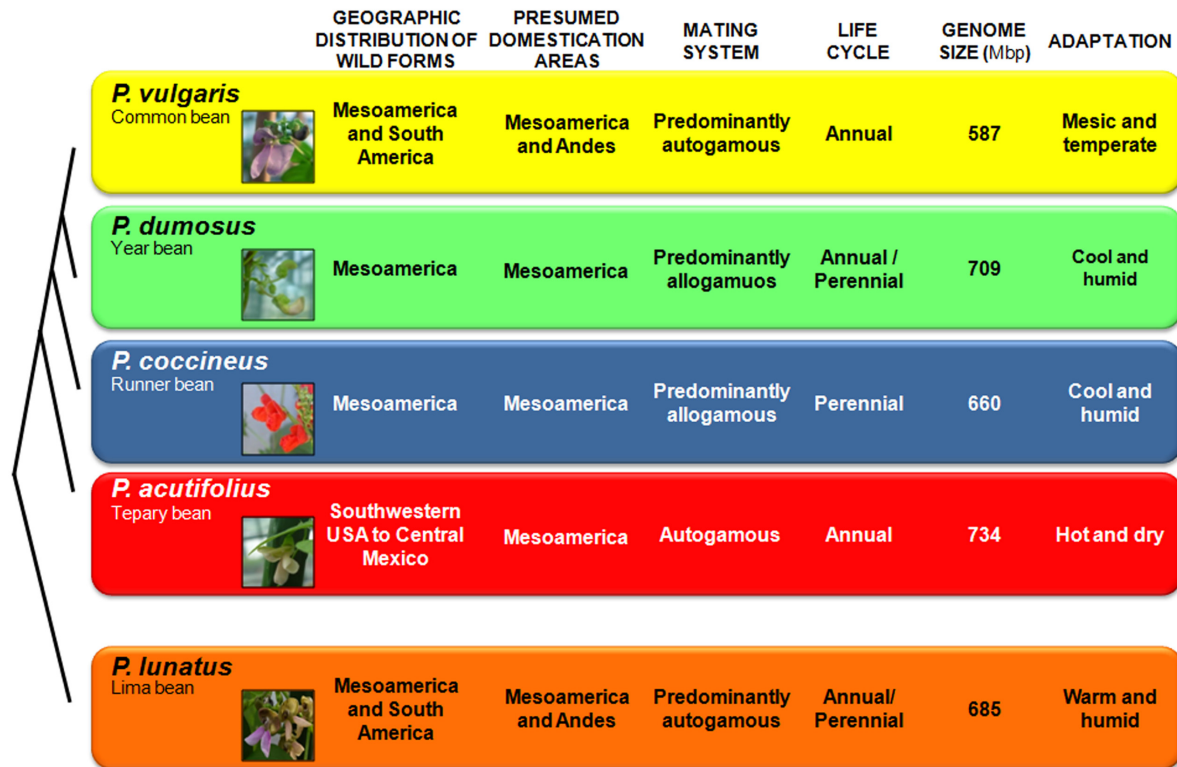


FIGURE 2 | Phylogenetic relationships of the five domesticated *Phaseolus* species, along with a comparison of their features. These relate to the geographic distribution of the wild forms, and their presumed domestication areas, mating systems, life cycles, genome sizes and ecological adaptations. Genome size data are from the Bennett and Leitch Plant DNA C-values database (release 6.0, Dec. 2012, <http://data.kew.org/cvalues/>).

distribution that is restricted to Mesoamerica (Figure 3). There have been numerous studies that have investigated the origin and evolution of *P. vulgaris*, which among these five domesticated *Phaseolus* spp. has the greatest economic importance. In contrast, there have been very few studies into the origins of *P. lunatus*, *P. acutifolius*, *P. coccineus*, and *P. dumosus*.

Phaseolus vulgaris

As proposed by Gepts (1988), a gene pool structure is identified by the observation within a biological species of strong population differentiation due to reproductive isolation, often associated to different adaptation, which can be due to geography and/or sexual incompatibility. Thus, the wild forms of the common bean, which grow from northern Mexico to north-western Argentina (Toro et al., 1990; Figure 3), are characterized by three eco-geographic gene pools. Two of these, the Andean and Mesoamerican, are the major gene pools of the species, and they include both wild and domesticated forms (Bitocchi et al., 2013). The third gene pool is represented by wild populations that grow in northern Peru and Ecuador, in a narrow altitudinal fringe on the western and eastern slopes of the Cordillera, a region characterized by diverse environmental conditions that differ from those on which the other wild Andean forms, including Colombian populations, grow (Debouck et al., 1993). Several studies have indicated the specific patterns of allelic

frequencies and linkage disequilibrium that are characteristic of the populations from northern Peru and Ecuador (Papa and Gepts, 2003; Kwak and Gepts, 2009; Nanni et al., 2011; Bitocchi et al., 2012; Desiderio et al., 2013; Rodriguez et al., 2016; Rendón-Anaya et al., 2017). This gene pool has only been described for wild populations, with no domesticated forms ever found. Moreover, the populations from northern Peru and Ecuador are characterized by a specific phaseolin type, known as 'Inca' (I), which is not found in individuals outside of this geographic location (Kami et al., 1995). In particular, the findings of Kami et al. (1995) indicated that the Inca sequence of the portion of the gene that codes for the protein phaseolin is ancestral to the other types of phaseolin in individuals from the Mesoamerican and Andean major gene pools. They thus indicated northern Peru and Ecuador as the area of origin of the common bean, from where, subsequently, the species became widespread northwards (Mesoamerica) and southwards (Andes), hence leading to the formation of the two major gene pools that are characteristic of *P. vulgaris*.

However, more recently, this evolutionary scenario was called into question by the studies of Rossi et al. (2009), and in particular, of Bitocchi et al. (2012), where their data analysis clearly indicated that both the Andean and the Inca gene pools are derived from independent introductions from Mesoamerica (Figure 4). In contrast to the previous studies that were based on

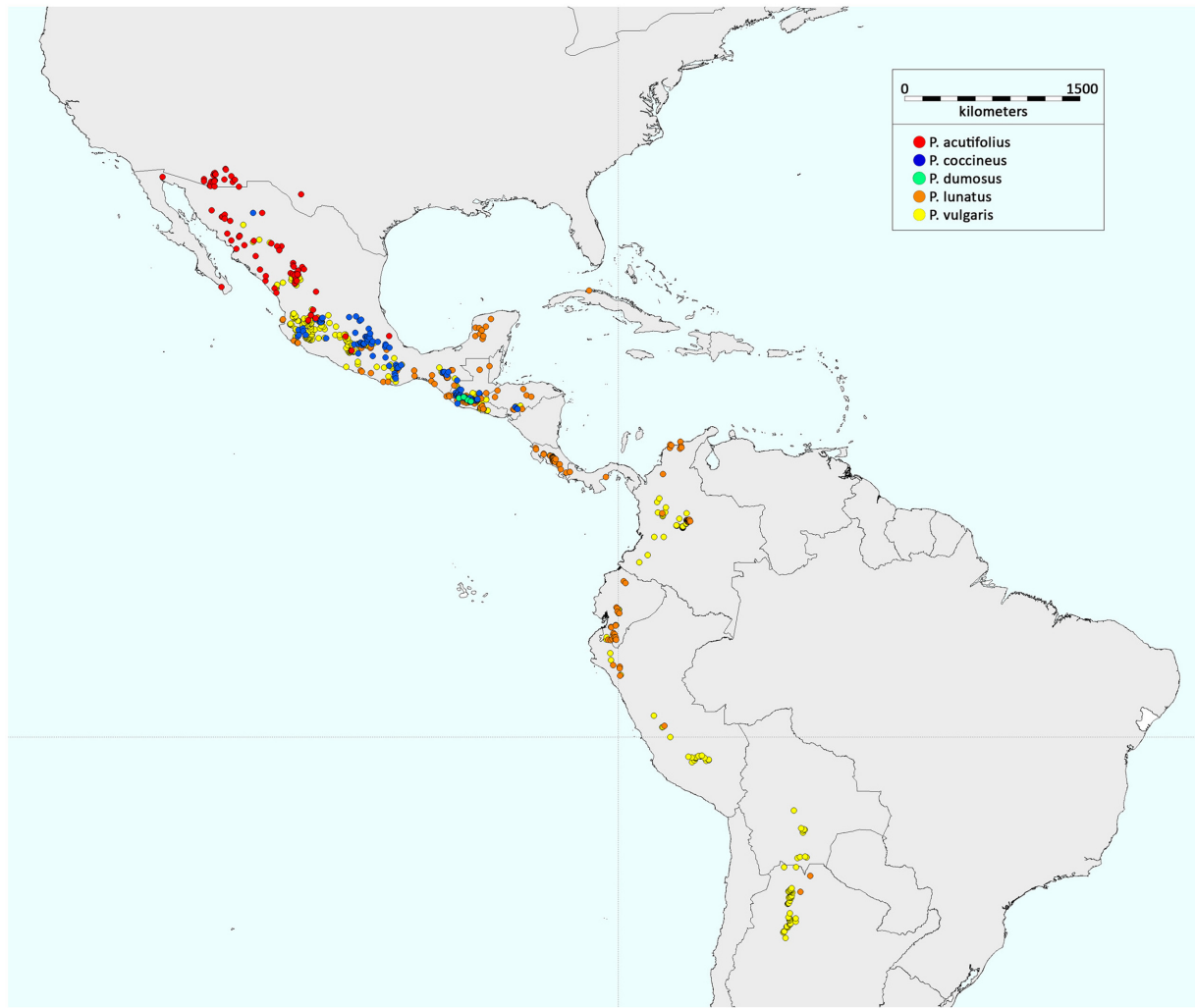


FIGURE 3 | Geographic distribution of the wild forms of the five domesticated *Phaseolus* species. The distributions were obtained by considering all of the wild accessions with passport data in the database of the International Center for Tropical Agriculture (CIAT).

multilocus molecular markers, Bitocchi et al. (2012) investigated the evolutionary history of the common bean using nucleotide sequences for five gene fragments. The first outcome was the clear confirmation of a bottleneck that occurred prior to domestication for the Andean gene pool. This was suggested previously by other studies (i.e., Rossi et al., 2009), although it was the lower mutation rate characteristic of nucleotide data compared to multilocus molecular markers that allowed the strong effect of the bottleneck on the genetic diversity of the Andean wild germplasm to be highlighted (i.e., a 90% reduction in diversity compared to the Mesoamerican wild gene pool; Bitocchi et al., 2012). Moreover, as sequence data from a single locus of a few 100 base pairs are less prone to recombination than multilocus molecular markers, these data allowed the identification of a strong population structure in Mesoamerica, although for the first time without any clear distinction between the Mesoamerican and Andean wild gene pools (Bitocchi et al., 2012). Indeed, the

phylogenetic relationships between the different groups showed two Mesoamerican groups (from Mexico) that appeared to be more closely related, one to the northern Peru–Ecuador gene pool and the other to the Andean gene pool. This led to the conclusion that each of these gene pools from South America originated through different migrations from the Mesoamerican populations of central Mexico. This hypothesis was also supported by subsequent studies (Schmutz et al., 2014; Rendón-Anaya et al., 2017). However, Rendón-Anaya et al. (2017), differently from Bitocchi et al. (2012) and Desiderio et al. (2013), by analyzing whole genome sequencing data from 18 *P. vulgaris* accessions (eight wild and two domesticated Mesoamerican accessions; one wild and two domesticated Andean accessions and five accessions from northern Peru and Ecuador), suggested that the northern Peru and Ecuador group should be considered as a sister species on the basis of the observation of a complete separation between the northern Peru and Ecuador genotypes



and a group composed by both Mesoamerican and Andean genotypes. However, from the works of Bitocchi et al. (2012) and Desiderio et al. (2013), presenting a larger sample compared to Rendón-Anaya et al. (2017), for both nuclear and chloroplast genome, wild populations from Mesoamerica closely related to the northern Peru and Ecuador group were identified, suggesting caution on the claim that the northern Peru and Ecuador group represents a different species. The whole genome sequencing analysis conducted by Schmutz et al. (2014), who estimated the divergence time between the Andean and Mesoamerican gene pools by applying demographic modeling, suggests that the wild populations in the Andes were derived from an ancestral Mesoamerican population ~165,000 years ago. The first attempt to date the split between the Andean and Mesoamerican gene pools was represented by the study of Gepts et al. (1999), where this event was estimated to have occurred ~500,000 years ago on the basis of the analysis of the α -amylase inhibitor and internal transcribed spacer sequence data. Mamidi et al. (2013) identified an earlier date (~110,000 years ago) compared to that estimated by Schmutz et al. (2014). This recent divergence is in agreement with the high similarity between the genomes and with the recent observation by Vlasova et al. (2016) that most of the bean-specific gene family expansion predates the split between Mesoamerica and the Andes.

The Other *Phaseolus* Crop Species

The *P. lunatus* wild forms are widely distributed from central Mexico to northern Argentina (Allard, 1960; Heiser, 1965;

Freytag and Debouck, 2002; **Figure 3**). Studies of the evolutionary history of *P. lunatus* have essentially been based on limited genomic data. In particular, most of these have relied on the analysis of two intergenic spacers of the chloroplast DNA (i.e., *atpB-rbcL*, *trnL-trnF*) and the sequence of the nuclear ribosomal 5.8S and flanking internal transcribed spacers (the ITS region) (Motta-Aldana et al., 2010; Serrano-Serrano et al., 2010, 2012; Andueza-Noh et al., 2013). Serrano-Serrano et al. (2010) analyzed these nuclear and non-coding chloroplast DNA markers in a collection of 59 wild Lima bean accessions and six related *Phaseolus* spp., three of which were of Andean distribution (i.e., *P. augusti*, *P. pachyrrhizoides*, and *P. bolivianus*), with the others distributed in Mesoamerica (i.e., *P. leptostachyus*, *P. marechalii*, and *P. novoleonensis*). Using neighbor-joining tree analysis, they identified three divergent wild Lima bean gene pools: Ecuador and northern Peru (AI); Mexico, and mainly the area to the west and northwest of the Isthmus of Tehuantepec (MI); and Mexico, and mainly the area to the east and southeast of the Isthmus of Tehuantepec (MII). Moreover, they suggested an Andean origin for wild Lima bean, as had been reported previously (Caicedo et al., 1999; Delgado-Salinas et al., 1999; Fofana et al., 1999). Their main evidence was the close phylogenetic relationship between the wild Lima bean and the related Andean *Phaseolus* species. In agreement with the study of Delgado-Salinas et al. (2006), Serrano-Serrano et al. (2010) indicated a relatively recent origin of *P. lunatus*, to during the Pleistocene and after the major Andean orogeny, ~2 to 5 Ma ago (Gregory-Wodzicki, 2000; Young et al., 2002). More recent studies based on the same markers and large samples of wild materials (Serrano-Serrano et al., 2012; Andueza-Noh et al., 2013) have confirmed the structure of the Mesoamerican wild populations, with the identification of the two main groups (MI, MII) with a geographic distribution that is probably related to adaptation to the different environments. Indeed, as indicated by Serrano-Serrano et al. (2012), the MI gene pool is mainly distributed in tropical dry forests over the Pacific coastal plain in Mexico at an average altitude of ~450 m a.s.l., with a small group of accessions on the western side of the Neo-Volcanic Axis at higher altitudes (from 1,250 to 1,810 m a.s.l.), while the MII gene pool is present in the Mexican lowlands (~550 m a.s.l.) along the Atlantic coast (Mexican gulf) and the Yucatan peninsula, and to the southeast of the Isthmus of Tehuantepec, the Caribbean, and South America.

Recently, Martínez-Castillo et al. (2014) analyzed 67 wild *P. lunatus* accessions from Mexico using 10 microsatellite (simple sequence repeats; SSR) markers. Through population structure analysis, they suggested that the genetic structure of the wild Lima bean in Mexico is more complex than previously thought, and they proposed three gene pools in Mesoamerica (MIa, MIb, and MII).

However, comparisons of the results of these studies show that there remain some disagreements, with different assignments to the diverse gene pools for the same accessions, according to the nuclear and chloroplast markers. Moreover, in some cases, the groups are not well supported statistically, and thus further studies are needed to more deeply understand the evolution of the Lima bean *P. lunatus*.

As mentioned above, the other three *Phaseolus* crop species (i.e., *P. acutifolius*, *P. coccineus*, and *P. dumosus*) are distributed in North and Central America. The wild forms of the tepary bean, which include *P. acutifolius* var. *acutifolius* and *P. acutifolius* var. *tenuifolius*, grow in the region from Central Mexico to southwestern USA (Blair et al., 2002; **Figure 3**). Indeed, this species is believed to have originated within this geographic range (Freeman, 1912, 1913; Nabhan and Felger, 1978; Manshardt and Waines, 1983).

The wild forms of the runner bean *P. coccineus* are distributed from Chihuahua (Mexico) to Panama (Delgado Salinas, 1988). The few archeological remains available (Delgado Salinas, 1988; Kaplan and Lynch, 1999) also indicate this area as the center of origin of *P. coccineus*.

For the wild form of the year bean *P. dumosus*, the geographic distribution is centered on a very narrow area in Guatemala, and this is where it has been suggested that *P. dumosus* originated (Schmit and Deboucq, 1991).

Adaptation

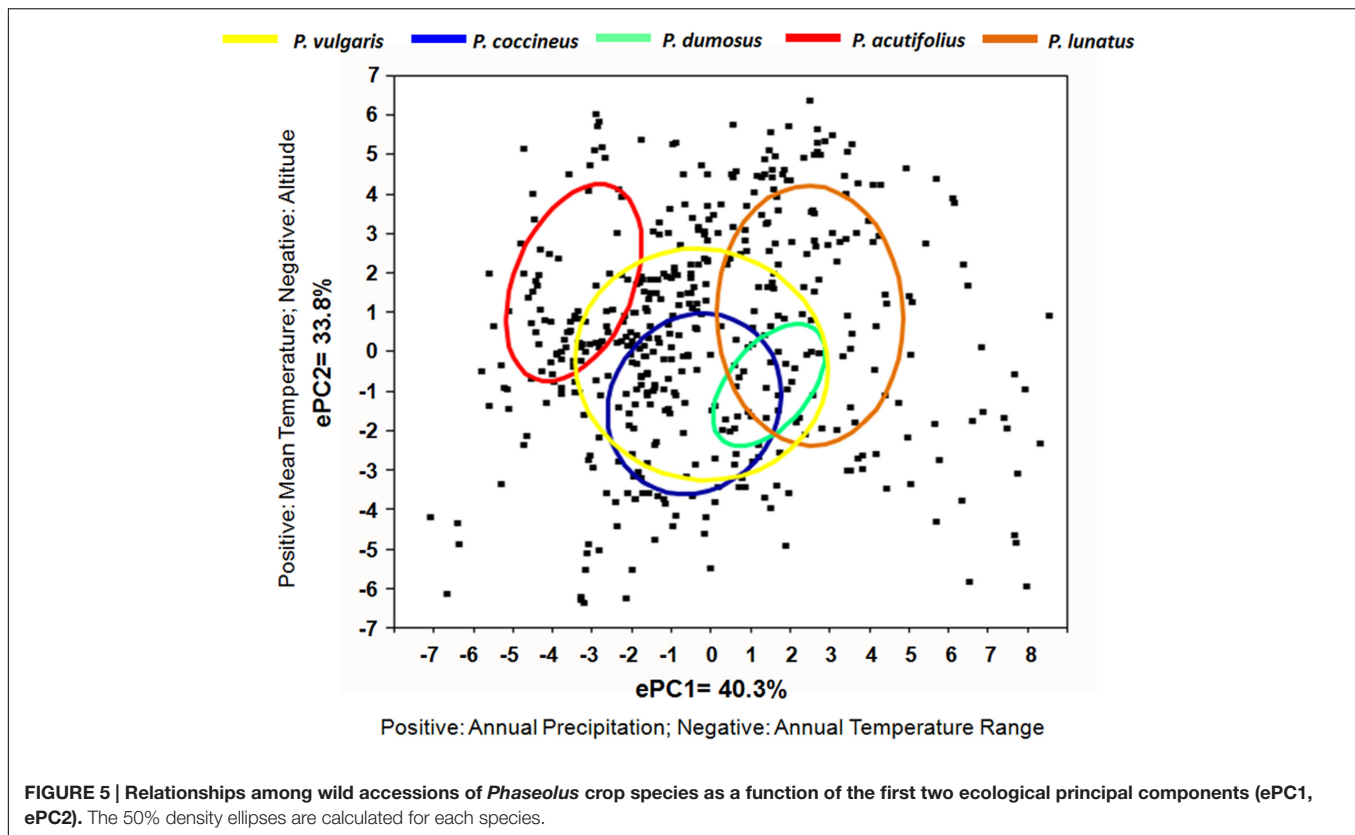
Genetic diversity is not uniformly distributed throughout the variety of climatic and environmental conditions under which a species grows. Consequently, it depends not only on demographic processes (e.g., genetic drift), but also on adaptation, in terms of the ability to cope with specific environmental conditions. These conditions can include extremes of cold and heat, lack or excess of water, different light intensities and duration, and diverse soil conditions and pest and disease pressures. As indicated above, the wild forms of *Phaseolus* crop species grow under a variety of different environmental and climatic conditions, and their geographic distribution mirrors their diverse patterns of adaptation to the different ecological niches, as well as their life histories and reproductive systems (**Figure 2**). *P. vulgaris* is adapted to warmer temperatures (mesic and temperate soil temperatures) at lower altitudes, with a rainfall of ~1,100 mm/year (Supplementary Table S1). *P. coccineus* is found in more humid environments, at cooler temperatures, and at higher altitudes, while *P. dumosus* is characterized by intermediate adaptation. *P. acutifolius* is a drought-tolerant species that originated in warmer and more arid environments (e.g., it is grown in the arid lands of Mexico and in southeast USA). Finally, *P. lunatus* is particularly suited to low-altitude humid and sub-humid climates, as well as warm temperate zones (**Figure 2** and Supplementary Table S1). These observations are confirmed by principal component analysis (**Figure 5**). This was calculated from the data on 20 ecological variables (Supplementary Table S1) using DIVA-GIS 7.5¹, as inferred from the geographical coordinates of the collection sites of the wild accessions for the passport data present in the database of the International Centre for Tropical Agriculture (CIAT). Moreover, as can be seen in **Figure 5**, among these five species, *P. vulgaris* shows the widest ecological adaptation.

There is little in the literature that is aimed at highlighting the genetic basis of adaptation, especially for the wild forms of these crop species, although a few recent studies have been

focused on *P. vulgaris*. Rodriguez et al. (2016) carried out a study on the environmental adaptation of wild *P. vulgaris*. They investigated the role of demographic processes (e.g., genetic drift) and selection for adaptation in the shaping of the current genetic structure of wild *P. vulgaris*. This analysis was based on 131 single nucleotide polymorphisms (SNPs) on a sample of 417 wild *P. vulgaris* accessions that are representative of the geographic distribution of the wild forms of the species. They first investigated the spatial distribution of the genetic diversity of the wild forms of *P. vulgaris*, using a landscape genetics approach that was based on an individual centered analysis to avoid sampling bias (**Figure 6**). Briefly, they delineated a circular neighborhood of 100-km radius around each georeferenced accession, and the accessions falling within each neighborhood were used to calculate the relative unbiased gene diversity, *He* (Nei, 1978). The correlation between *He* and the neighborhood size was not significant ($r = 0.040$, $n = 299$, $P = 0.482$), and the mean size of each neighborhood was 40.6 individuals and 83.3% of the neighborhoods included >10 individuals (Rodriguez et al., 2016). By using this approach, they observed high genetic diversity across Mexico, from the state of Oaxaca to Durango, with a depression in the genetic diversity in an area that lies approximately across the regions of Guerrero, Morelos, Puebla and Estado de Mexico. Possible explanations for the reduced diversity of this area in Mexico include natural selection due to an environment that is too arid for *P. vulgaris*, or a genetic bottleneck caused by the volcanic activities that were frequent in this area in ancient times (Márquez et al., 1999; Siebe, 2000; Siebe et al., 2004; Plunket and Uruñuela, 2012). Low diversity was also found in Guatemala, Costa Rica and Colombia, and particularly in the Honduras (**Figure 6**).

The approach applied by Rodriguez et al. (2016) also highlighted local adaptation at the continental level. The rationale behind the study was that the use of spatial data in combination with genetic diversity data would allow discrimination between the effects of geography and ecology; i.e., demographic processes vs. selection (Bradburd et al., 2013; Wang et al., 2013; Guillot et al., 2014; Kraft et al., 2014). As evidence of the effectiveness of this approach, it was possible to disentangle the effects of geography from ecology in the shaping of the genetic patterns observed, and correlations between markers and ecological variables were detected. By scanning the SNP markers used for the analyses, 26 loci (19.8%) were identified as under signatures of selection, seven of which (5.3%) showed strong probability levels. Although the proportion of loci under selection might be overestimated, as they were chosen among genes that were putatively involved in adaptation, different loci were found to have compatible functions with adaptation features, such as cold acclimation, chilling susceptibility, and mechanisms related to drought stress. Moreover, as well as the well-delineated genetic groups in the Mesoamerican gene pool, they demonstrated global structures for both the neutral loci and the loci under selection. Overall, these data suggest that the origin of the geographic structures might be the outcome of the expansion of the species and gene flow, including crop-to-wild introgression (Papa and Gepts, 2003; Papa et al., 2005), as was

¹<http://www.diva-gis.org/>



also recently confirmed by Rendón-Anaya et al. (2017) using whole genome sequencing analysis. Nonetheless, the adaptation to different environmental conditions might have led to the present population structure, with the subsequent limited long-range gene flow and divergent selection along an ecological cline of variation.

DOMESTICATION OF *PHASEOLUS* SPP.

In the context of domestication, the *Phaseolus* spp. represent a unique model, where these five closely related species provide a unique example of multiple parallel domestication events, which allow to investigate this evolutionary process as a kind of replicate experiment. Two of these species (i.e., *P. vulgaris* and *P. lunatus*) have gone through at least two isolated and independent domestication processes, due to the distribution of the wild forms. For the common bean *P. vulgaris*, two independent events in the Americas (in Mesoamerica and the Andes) have been documented in several studies (for review, see Bellucci et al., 2014b), where the two major domesticated gene pools originated (Bitocchi et al., 2013). A similar scenario has been observed for *P. lunatus*, where at least two independent domestication events have been suggested. One of these was in the Andes, which appears to have given rise to the large-seeded landraces collectively known as the ‘big Lima’ cultivars (Motta-Aldana et al., 2010). The other was in Mesoamerica, which appears to have given rise to the great variety of small-seeded

Mesoamerican landraces (Motta-Aldana et al., 2010; Serrano-Serrano et al., 2012).

Not so much is known for the other three domesticated *Phaseolus* spp.. Through an analysis of a small set of wild and domesticated *P. coccineus* accessions and using chloroplast SSRs, Angioi et al. (2009a) identified two different wild genetic groups that paralleled the differentiation between two groups of the domesticated accessions, which suggested multiple domestication events for *P. coccineus* in Mesoamerica. A single domestication event was suggested for *P. dumosus* by Schmit and Debouck (1991) on the basis of seed protein data that were analyzed on a sample of 163 wild and domesticated accessions, as also for *P. acutifolius*, through studies based on phaseolin (Schinkel and Gepts, 1988), isoenzymes (Garvin and Weeden, 1994), SSRs (Blair et al., 2012), and SNPs (Gujaria-Verma et al., 2016). In particular, Gujaria-Verma et al. (2016) analyzed 645 SNPs markers on a wide sample that included both wild and domesticated *P. acutifolius* accessions, and they reported that domesticated tepary beans formed a tightly linked cluster that was subdivided into two major groups based on their eco-geographical origin (Central America and USA/Mexico). Moreover, the domesticated accessions were clearly separated from the wild, which suggested that it was likely that there had been an early domestication event that was followed by separation based on regions.

Overall, for the *Phaseolus* species, at least seven independent domestication events have occurred, with examples seen of both multiple and single domestication processes.

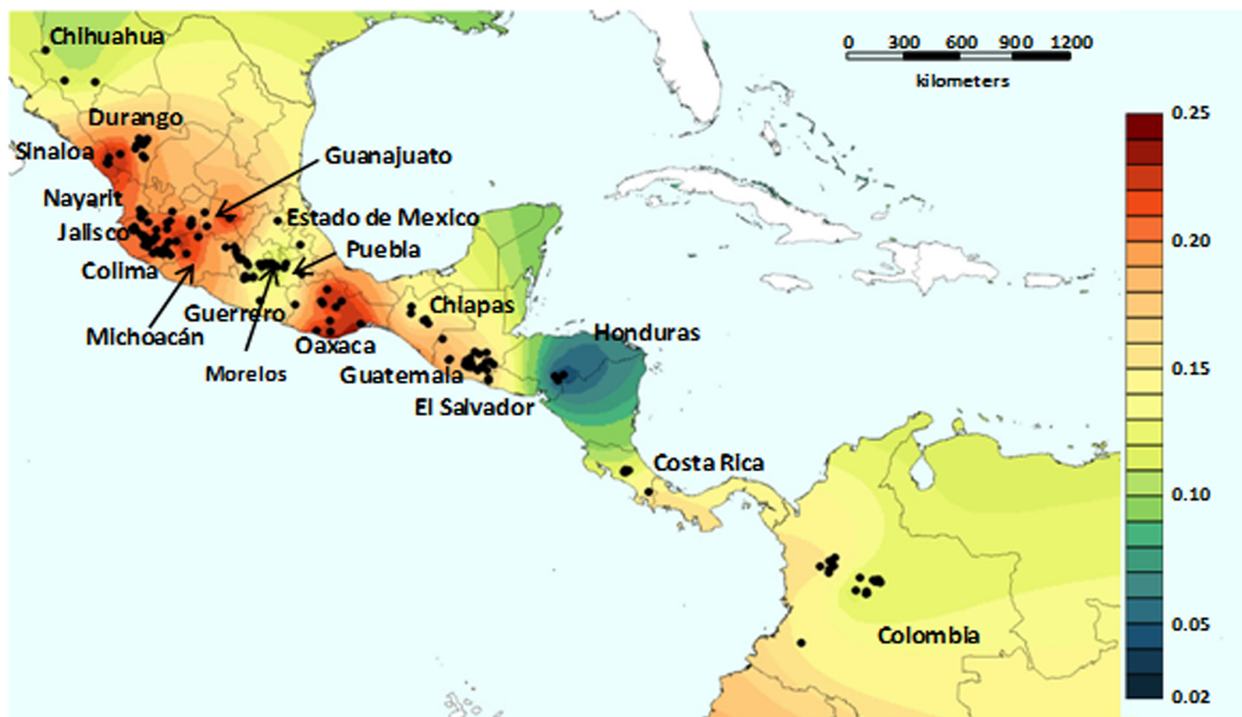


FIGURE 6 | Genetic heat-map generated by diversity interpolation based on an individual-centered approach. The genetic topography of *P. vulgaris* varies from low (blue) to high (red) diversity levels, which indicate the generally higher levels of diversity in Mesoamerica with respect to the Andes (from Rodríguez et al., 2016; permissions for reproduction have been obtained from the copyright holders).

Geographic Areas of *Phaseolus* ssp. Domestication

One of the major issues of evolutionary studies that are focused on the domestication of plant species is to identify the geographic areas where they were domesticated (Harlan, 1975).

Phaseolus vulgaris

Considering the process of domestication within each gene pool, after the long debate for *P. vulgaris* on this issue (for review, see Bellucci et al., 2014b), recent studies have indicated that a single domestication event occurred within each gene pool (Mamidi et al., 2011; Nanni et al., 2011; Bitocchi et al., 2013).

For the Mesoamerican gene pool, several important crops were domesticated in Mexico, including maize, squash and common bean, as has been well documented in different studies based on both archeological and molecular data (for review, see Bitocchi et al., 2013). Recently, based on SSR data, Kwak et al. (2009) suggested the Rio Lerma–Rio Grande de Santiago basin in western-central Mexico as the putative geographic area where the common bean *P. vulgaris* was domesticated. The domestication area they suggested for the common bean did not overlap with that indicated for maize (central Balsas River drainage; Matsuoka et al., 2002; van Heerwaarden et al., 2011). Thus Kwak et al. (2009) proposed that maize and the common bean were probably domesticated in different regions and were instead reunited later in a single cropping system: the milpa, the cropping system that forms the basis of the Mesoamerican traditional

agriculture based on the intercropping of maize (*Zea mays* L.), the common bean, and squash (*Cucurbita* spp.). Bitocchi et al. (2013) used nucleotide data to suggest a different location for the domestication of the common bean: Oaxaca Valley, in the south of Mexico. This area does not coincide with that proposed by Kwak et al. (2009) for the common bean and that of maize (Matsuoka et al., 2002); however, it overlaps with one of the first areas in the spread of maize through human migration, along the Mexican rivers (Zizumbo-Villarreal and Colunga-GarcíaMarín, 2010).

The pinpointing of the geographic area where domestication took place for the common bean in the Andes has been more difficult, due to the low diversity characteristic of this germplasm. However, different studies have tried to identify the geographic area for this domestication. Chacón et al. (2005) analyzed the polymorphism at the level of the chloroplast DNA for a wide sample of common bean accessions from South America, and they suggested that central-southern Peru was the cradle of domestication of the Andean gene pool. Beebe et al. (2001) used amplified fragment length polymorphisms (AFLPs) to define a strict relationship between the domesticated forms and the wild beans from eastern Bolivia and northern Argentina, and suggested this location as the putative area of domestication, which was also supported recently by the study of Bitocchi et al. (2013).

More recently, on the basis of SNP data, the locations proposed by Bitocchi et al. (2013) for both the Mesoamerican and

Andean gene pools were confirmed by the study of Rodríguez et al. (2016). Here, Rodríguez et al. (2016) integrated the results obtained from spatial, phenotypic and molecular data with those from different disciplines, including archeological and glotto-chronological data, to pinpoint the domestication sites in Mesoamerica and the Andes. The low genetic distances between the wild forms and the domesticated forms indicated these as genetic groups located in the Oaxaca Valley in Mesoamerica and in a region from northern Argentina to southern Bolivia in the Andes, respectively. Consistent with these data, previous archeological data have indicated the early occurrence of domestication in these areas (Tarrago, 1980; Kaplan and Lynch, 1999). Recent glotto-chronological studies have also supported these conclusions, as within these areas the homeland sites of proto-languages for which ancient bean words were reconstructed showed times that are compatible with the domestication of *P. vulgaris* (Brown et al., 2014). We expect that future studies will better refine the areas of domestication. For this purpose will be very important also to conduct appropriate explorations. Indeed, as shown by Zizumbo-Villarreal et al. (2009) additional wild bean populations, not yet included in germplasm banks, can still be identified.

The Other *Phaseolus* Crop Species

Similar to the common bean, the Lima bean *P. lunatus* has been domesticated at least twice in the Americas: once in the Andean region, and at another time in Mesoamerica. However, further studies are needed to investigate the domestication process of this species more deeply. Indeed, this still appears unclear especially for the Mesoamerican gene pool, where there remains open debate concerning its single or multiple domestication (Motta-Aldana et al., 2010; Serrano-Serrano et al., 2012; Andueza-Noh et al., 2013).

Motta-Aldana et al. (2010) analyzed chloroplast DNA and ITS polymorphisms in a sample of wild and landrace accessions of *P. lunatus*, through which they suggested one domestication event in the Andes of northwestern Peru and southern Ecuador, and a second in central-western Mexico, which they indicated as more likely to be in the area to the north and northwest of the Isthmus of Tehuantepec. Consistent with this study by Motta-Aldana et al. (2010), in their analysis of ITS data on a large sample of Mesoamerican wild and domesticated *P. lunatus*, Serrano-Serrano et al. (2012) proposed a single event of domestication in Mesoamerica. As indicated above, they showed evidence for two wild Mesoamerican gene pools with mostly contrasting geographic distributions. In their cluster analysis, all of the Mesoamerican landraces clustered together with the wild accessions from the MI gene pool, which is characteristic of central-western Mexico. This suggests a unique domestication event in an area of the states of Nayarit–Jalisco or Guerrero–Oaxaca, and not on the Peninsula of Yucatan, where *P. lunatus* is currently widespread and diverse.

Andueza-Noh et al. (2013) used two intergenic spacers of chloroplast DNA to confirm these Mesoamerican and Andean gene pools for *P. lunatus* (Gutiérrez-Salgado et al., 1995; Maquet et al., 1997; Lioi et al., 1998; Fofana et al., 2001; Motta-Aldana et al., 2010), and the two genetically and geographically distinct

groups within the Mesoamerican gene pool (MI, MII; Motta-Aldana et al., 2010; Serrano-Serrano et al., 2010, 2012). They pinpointed the domestication area for the Andean gene pool as mid-altitude western valleys between Peru and Ecuador in South America, as had already been suggested (Gutiérrez-Salgado et al., 1995; Fofana et al., 2001; Motta-Aldana et al., 2010). However, in contrast to the previous studies, Andueza-Noh et al. (2013, 2015) indicated multiple origins of domestication in Mesoamerica for *P. lunatus*. For the MI group, they indicated western-central Mexico as the domestication area, while they proposed a more restricted geographic area between Guatemala and Costa Rica for the MII group. However, they were aware that more studies involving more comprehensive geographic and genomic sampling are needed to define how the domestication processes and gene flow have shaped the current genetic structure of *P. lunatus* landraces.

For the other three *Phaseolus* crops (i.e., *P. acutifolius*, *P. coccineus*, and *P. dumosus*), nothing much is known about their domestication or where this process might have taken place. There have been very few studies on *P. acutifolius* and *P. coccineus*. For the tepary bean *P. acutifolius*, early studies based on phaseolin and isozyme analysis highlighted the controversy over the number of domestication events. Here, some studies proposed two domestication events in the northern and southern parts of the range (Manhardt and Waines, 1983), and others suggested a single origin but different locations for the domestication, as either in the Mexican state of Durango (Schinkel and Gepts, 1988) or the states of Sinaloa or Jalisco (Garvin and Weeden, 1994). The more recent study of Blair et al. (2012) was based on SSR data on a wide sample of wild and domesticated *P. acutifolius* accessions from its area of distribution. They indicated, as mentioned above, that a single domestication event was likely, and that the cultivars were most closely related to *P. acutifolius* var. *acutifolius* accessions from Sinaloa and northern Mexico.

Phaseolus coccineus is native to Mexico, Guatemala and Honduras (Delgado Salinas, 1988), and the wild forms are probably not all ancestral to the cultivated form. However, the area(s) where the domestication of *P. coccineus* took place are still not known. Spataro et al. (2011) used SSR data with a collection of wild and domesticated accessions, and they showed that most of the Mesoamerican landraces they examined closely resembled wild genotypes from Guatemala and Honduras, while only a few resembled wild Mexican forms. This would suggest that *P. coccineus* domestication either took place in that area, or that two domestication events took place (in Guatemala–Honduras and Mexico, separately) followed by extensive hybridisation with the cultivated forms from Guatemala and Honduras.

The distribution of wild *P. dumosus* (the year bean) is extremely narrow on the basis of findings to date, and it appears to be concentrated only in central southwestern Guatemala. Schmit and Debouck (1991) used phaseolin data together with information on vernacular names, and they reported that there is a single gene pool that was domesticated from a wild ancestor that is still present in Guatemala. Thus, they indicated a single domestication in Guatemala, and subsequent diffusion toward the humid highlands of Chiapas, Oaxaca, Puebla and

Veracruz in Mexico, and toward Costa Rica and the northern Andes.

Domestication Bottleneck

The population genetics model of domestication predicts a reduction in diversity and increased divergence between wild and domesticated populations, due to demographic factors that affect the whole genome, and because of selection at target loci. Several interesting insights can be revealed by comparisons between different species (Glémin and Bataillon, 2009). There are many examples in the literature that have used different molecular markers and nucleotide data to show the reduction in genetic diversity of crop species compared with their wild progenitors (for review, see Bitocchi et al., 2013). Allogamous species, such as maize (*Z. mays*), are generally characterized by lower genetic bottleneck effects compared to autogamous species, such as the common bean *P. vulgaris*, even if other factors can have relevant roles, such as the life history (Bitocchi et al., 2013). Resequencing data have confirmed that in autogamous species, such as soybean (*Glycine max*) and rice (*Oryza sativa*, variety *japonica*) (Lam et al., 2010; Xu et al., 2012), reductions in diversity have arisen due to domestication, as also reported for silkworm and for mammalian species (Xia et al., 2009; vonHoldt et al., 2010; Lippold et al., 2011).

Phaseolus vulgaris

For the common bean *P. vulgaris*, different studies have clearly identified a bottleneck due to domestication in both the Mesoamerican and Andean gene pools (e.g., Papa et al., 2005; Kwak and Gepts, 2009; Rossi et al., 2009; Mamidi et al., 2011; Nanni et al., 2011; Bitocchi et al., 2013, 2016a; Bellucci et al., 2014a). However, the reduction in diversity in the domesticated forms compared to the wild forms was greater in Mesoamerica compared to the Andes. Indeed, Bitocchi et al. (2013) reported that this loss of diversity was threefold greater for Mesoamerica compared to the Andes, and this was explained as the result of the bottleneck that occurred before domestication in the Andes. This thus strongly impoverished the genetic variability of the Andean wild germplasm, which led to a minor effect of the subsequent domestication bottleneck (i.e., sequential bottleneck). These outcomes demonstrate that the understanding of the level and the structure of genetic diversity of a species needs to be accompanied by a close appraisal of its evolutionary history.

Bellucci et al. (2014a) exploited next-generation sequencing technologies to analyze changes at the transcriptome level in *P. vulgaris* accessions from Mesoamerica, to investigate the domestication process in this gene pool more deeply. They used RNA sequencing technology and *de novo* transcriptome assembly to compare representative sets of wild and domesticated accessions of the common bean from Mesoamerica, and they reported the profound effects that domestication imposed on the genome variation and gene expression patterns of the common bean. Indeed, they showed that in addition to reduced nucleotide variation, the domesticated common bean showed reduced gene expression diversity, while in maize, the same reduction was not seen in parallel with reduced effects of domestication for nucleotide diversity (Hufford et al., 2012; Swanson-Wagner

et al., 2012). The expressed genomic regions lost half of the wild-bean nucleotide diversity during the domestication in Mesoamerica, and in parallel, the effects of domestication significantly decreased the diversity of gene expression (by 18%). For the first time, this demonstrated that loss of genetic variation has direct genome-wide phenotypic consequences on transcriptome diversity. The contigs identified as differentially expressed (in the comparison of domesticated vs. wild) were mostly down-regulated in the domesticated forms (by 74%). This indicated loss-of-function mutations (which are relatively frequent compared to gain-of-function changes) as a largely available source of variation that supports selection during rapid environmental changes (Olson, 1999). Such was the case for the transition from the wild to the cultivated agro-ecosystems. In support of this, as first noted by Darwin (1859), in domesticated plants, the domestication traits have a recessive genetic nature (Lester, 1989).

In addition to the case of differentially expressed genes, the genome-wide gene expression reported by Bellucci et al. (2014a) for the domesticated common bean *P. vulgaris* was on average lower than for the wild. They interpreted this result as the accumulation of slightly deleterious mutations due to hitchhiking (mostly loss-of-function, or with reduced expression) in *P. vulgaris*, and considered this as the ‘cost of domestication.’ This accumulation of loss-of-function (or reduced expression) mutations might also have been due to reduced effective recombination, which would have increased the frequency of deleterious mutations in the domesticated pool, and have had a negative influence on the fitness, as was suggested in rice (Lu et al., 2006).

The Other *Phaseolus* Crop Species

In their analysis of chloroplast DNA and ITS polymorphisms in a sample of wild and landrace accessions of *P. lunatus* (the Lima bean), Motta-Aldana et al. (2010) observed a severe reduction in genetic diversity because of domestication in both the Mesoamerican and Andean gene pools (the MI wild accessions were used for co-mutations); in particular, the loss of diversity appeared stronger according to chloroplast DNA data (100%, 92.1%, for the Mesoamerican and Andean gene pools, respectively) than for ITS data (46.6%, 58.5%, respectively). This was confirmed for the Mesoamerican gene pool by Serrano-Serrano et al. (2012) and Andueza-Noh et al. (2013, 2015), through analysis of two intergenic spacers of chloroplast DNA (loss of diversity, 60.83%), SSR markers (loss of diversity, 44%), and the ITS region of ribosomal DNA (loss of diversity, 53%).

A bottleneck of domestication was also seen for *P. acutifolius*. Genetic diversity within the domesticated forms of the tepary bean is low, as has been shown by studies of phaseolin patterns (Schinkel and Gepts, 1988), isozymes (Schinkel and Gepts, 1989; Garvin and Weeden, 1994), AFLPs (Muñoz et al., 2006), and SSR markers (Blair et al., 2012). Considering the year bean *P. dumosus*, Schmit and Debouck (1991) analyzed phaseolin data and showed that the wild ancestral forms in central Guatemala show the highest diversity.

A different scenario has been reported in the few studies on domestication of the runner bean *P. coccineus*. Escalante et al.

(1994) indicated that the domestication process did not erode the genetic diversity of *P. coccineus*, and that the similar levels of genetic variation among the wild and cultivated materials were mainly due to the high gene flow between these two forms. This result was confirmed by Spataro et al. (2011) through an analysis of SSR data on a sample of wild and domesticated runner bean accessions.

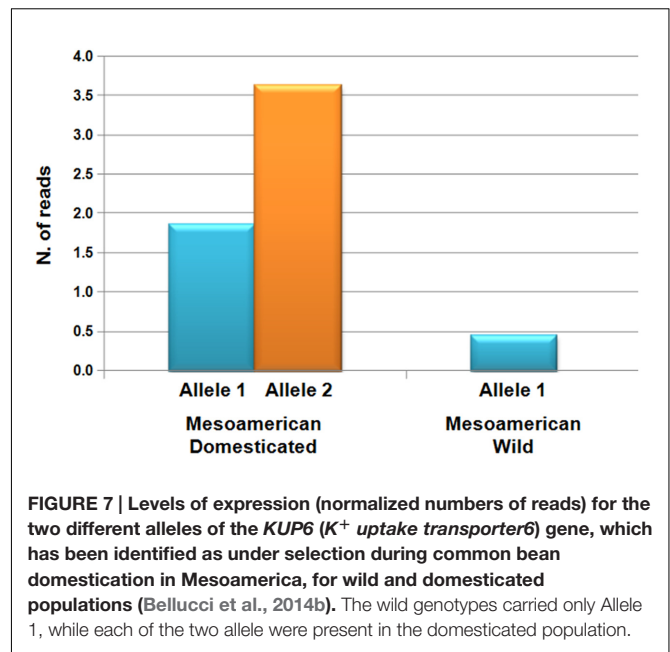
Signatures of Selection during Domestication

Identification of the genes involved in the domestication process and knowledge of the regions of the genome where those genes are located and of the proportion of the genome affected by domestication are key to better exploitation of the diversity present in the wild relatives, and to enhance the achievements in breeding and crop improvement (Tanksley and McCouch, 1997; McCouch, 2004). As proposed by Cavalli-Sforza (1966) and Lewontin and Krakauer (1973), the identification of loci involved in adaptive processes can be obtained from population genetics expectations that predict that while drift has a homogeneous effect over the genome, selection is acting only for target loci and related linked loci due to the lack of recombination (hitchhiking). Thus, selected (and linked loci) are expected to depart from neutral expectation of diversity and divergence parameters. Moreover, as proposed by Papa et al. (2005), in gene flow between crops and wild forms, aberrant patterns of divergence and diversity can also be determined by the combined actions of asymmetric migration (Papa and Gepts, 2003) and selection at target loci.

Papa et al. (2007) used 2,506 AFLPs for a whole genome scan for the signature of selection due to domestication, and they estimated that about 16% of the genome of the common bean *P. vulgaris* appeared to be under the effects of selection. Bellucci et al. (2014a) used RNA sequencing technology, and after simulating the demographic dynamics during domestication, they reported that 9% of the genes were actively selected during domestication in Mesoamerica. Furthermore, in these contigs, selection induced a further reduction in the diversity of gene expression (by 26%), and was associated with a fivefold enrichment of the differentially expressed genes.

Bellucci et al. (2014a) also carried out a survey on the function of a subset of contigs that are putatively under selection, to determine whether they are known to be associated with the domestication process in other species, using either direct experimentation or through their function. Interestingly, among the genes putatively under selection that showed greater genetic diversity in the wild compared with the domesticated form, they found sequence homologs to: (i) genes that are involved in 'light' response pathway; e.g., *GIGANTEA* (GI), which has a pivotal role in the photoperiodic response (Mizoguchi et al., 2005; Hecht et al., 2007; Kim et al., 2012); (ii) genes that are pivotal to ensure correct hormonal perception, transport or biosynthesis; (iii) genes that are involved in seed development and traits; and (iv) genes that are involved in responses to environmental stress.

Another interesting example is the homolog of *YABBY5* (*YAB5*), which is a transcription factor that is implicated in



the regulation of seed shattering in cereal species, including sorghum (*Sorghum bicolor*), rice and maize (Lin et al., 2012). In most cases, Bellucci et al. (2014a) found evidence of positive selection associated with domestication, but in a few cases, this selection had increased the nucleotide diversity in the domesticated pool at a target locus associated with abiotic stress responses, flowering time, and morphology. In particular, for 2.8% of the genes putatively identified to be under the effect of selection by Bellucci et al. (2014a), there was no diversity in the wild forms, while there was diversity in the domesticated. They explained this as due to novel mutations (or standing variations) that were selected because of the crop expansion into new environments (diversifying selection) with unexpected biotic and abiotic stress, or because of selection for traits that improved the use of the plant organs by humans (de Alencar Figueiredo et al., 2008). An interesting example was given by the functional analysis of the drought- and growth-related (Osakabe et al., 2013) *KUP6* (*K⁺ uptake transporter6*) gene, where this was significantly overexpressed in the domesticated compared to the wild (Figure 7), as if domestication had also increased the functional diversity of selected genes in addition to the increased nucleotide diversity.

Schmutz et al. (2014) also investigated whether their candidate genes were implicated in important domestication traits, such as flowering time and seed size. A total of 38 flowering genes were identified in the Mesoamerican and Andean candidate lists, while another subset of 15 genes was found to be associated with seed size in genome-wide association studies, and 11 genes contained SNPs that were associated with seed weight.

Recently, Bitocchi et al. (2016a) analyzed nucleotide sequences from a set of 49 gene fragments from a sample of 39 wild and domesticated Mesoamerican accessions of *P. vulgaris*. In this study, they applied the same approach as Bellucci et al. (2014a), and they identified several loci that showed signatures of

selection during common bean domestication in Mesoamerica. In particular, they had the possibility to see if their candidates were detected as outliers also in other studies of varying sizes, data types, and methodologies (Bellucci et al., 2014a; Schmutz et al., 2014; Rodriguez et al., 2016). They thus obtained independent evidence that four genes (i.e., *AN-Pv33*, *AN-Pv69*, *AN-DNAJ*, and *Leg223*) were targets of directional selection during common bean domestication. The gene function investigation for these genes highlighted that they are involved in plant resistance tolerance to biotic and abiotic stresses, such as heat, drought, and salinity. Moreover, another important outcome of Bitocchi et al. (2016a) was related to the observed excess of non-synonymous mutations in the domesticated germplasm. In particular, they observed a significantly higher frequency of polymorphisms in the coding regions compared to non-coding regions only in the domesticated beans. These mutations were mostly non-synonymous and were recently derived mutations present in genes related to responses to biotic and abiotic stresses. These data cannot be fully explained by the cost of domestication alone, but support a scenario where new functional mutations were selected for adaptation during domestication, showing that domestication also increased the functional diversity at target loci that enable the domesticated forms to successfully compete during the expansion and adaptation to new agro-ecological growing conditions.

Phenotypic Convergent Evolution

The key aspect of domestication is the convergent phenotypic evolution that is associated with the adaptation to a novel agro-ecosystem, and to human needs. For instance, most domesticated animals were selected to maximize the production of useful products (e.g., meat, milk, and wool) and for their docile behavior, while crops were selected for the size of the plant organs used by humans (e.g., seeds and fruit) and for reduced, or lack of, seed dispersal. For these reasons, domestication provides us with a unique tool to understand the process of adaptation, to test evolutionary hypotheses, and to identify the molecular basis of phenotypic diversity. Several interesting insights can be revealed by comparisons among different species (Glémin and Bataillon, 2009), where, for instance, the population genetics model of domestication predicts a reduction in diversity and increased divergence between wild and domesticated populations due to demographic factors that affect the whole genome and because of selection at target loci.

In this regard, an example is given by the study of Nanni et al. (2011), in which wild and domesticated *P. vulgaris* accessions were analyzed by sequencing a genomic region of ~1,200 bp (*PvSHP1*) that is homologous to *SHATTERPROOF-1* (*SHP1*), a gene involved in the control of fruit shattering in *Arabidopsis thaliana* (Liljegren et al., 2000). The loss of fruit shattering has been under selection in most seed crops, to facilitate seed harvesting (Purugganan and Fuller, 2009), while in wild plants, this feature is a fundamental trait enabling seed dispersal. Expressed sequences that correspond to *SHP1* have also been identified in other species, such as, tomato, where it was indicated as having an important role in the regulation of both fleshy fruit expansion and the ripening process (Vrebalov et al., 2009),

which are together necessary to promote seed dispersal of fleshy fruit. In legumes, sequences orthologous to *Arabidopsis SHP1* have been identified in *Medicago*, pea and soybean (Hecht et al., 2005). Nanni et al. (2011) mapped *PvSHP1* on linkage group Pv06 of the common bean genome, and showed that it did not co-segregate with the *St* locus, which is responsible for the presence or absence of pod string, and was mapped on linkage group Pv02 using a domesticated (Midas) × wild (G12873) RIL population by Koinange et al. (1996). These results suggested that *PvSHP1* is not responsible for the observed phenotypic variation in *P. vulgaris* for fruit shattering. Similar results were found by Gioia et al. (2013b), who sequenced and mapped *PvIND* on the common bean genome, a sequence that is homologous to the *INDEHISCENT* gene (*IND*), which is the primary factory required for silique shattering in *A. thaliana*. *PvIND* mapped near the *St* locus; however, the lack of complete co-segregation between *PvIND* and *St* and the lack of polymorphisms at the *PvIND* locus correlated with the dehiscent/ indehiscent phenotype suggested that *PvIND* is not directly involved in pod shattering and is not the gene underlying the *St* locus (Gioia et al., 2013b).

Recently, Murgia et al. (2017) developed a phenotyping approach in *P. vulgaris* to evaluate the shattering syndrome in a segregating population. This is a promising approach for the identification of genetic factors that control the shattering trait in the common bean, and it will greatly facilitate comparative studies among legume crops, and also gene tagging.

Within the same species, the study of Schmutz et al. (2014) represents the first example of the possibility to investigate convergent evolution between the two gene pools of the common bean *P. vulgaris*. Indeed, a comparison of the results of selection in the two gene pools, in which independent domestications occurred, allowed them to determine whether to obtain the same convergent phenotypes, evolution took part in the selection of the same genomic regions or of completely different set of genes that code for the same phenotypes. Interestingly, only a small portion of the genome and of genes identified as putatively under selection during domestication were shared between the two gene pools, which suggested different genetic routes to domestication (Schmutz et al., 2014). This outcome appears to suggest that the sexually compatible Mesoamerican and Andean lineages with similar morphologies and life cycles underwent independent selection upon distinct sets of genes. However, taking into account that explicit demographic modeling was not used to generate an expectation of the number of potential false positive regions by Schmutz et al. (2014), another possible explanation for this result is that the lack of correlation between the two gene pools is due to a high level of false positives; i.e., regions of the genome with reduced diversity due to the stochastic effects of domestication bottlenecks. Regarding this consideration, Bitocchi et al. (2016a) compared their candidates for selection during domestication of common bean in Mesoamerica with those of other studies (Bellucci et al., 2014a; Schmutz et al., 2014; Rodriguez et al., 2016), and found that two (*AN-Pv69*, *AN-DNAJ*) out of the four strong candidates identified were detected as outliers by Schmutz et al. (2014).

only during Andean domestication. This implies that more studies are needed either to support or refute the lack of correlations between the two gene pools found by Schmutz et al. (2014).

Dissemination of *Phaseolus* Crop Species Outside Their Centers of Origin

Beans are widely cultivated out of the Americas, and especially the common bean, which is the main *Phaseolus* crop species cultivated worldwide. For this reason, almost all the literature that has focused on investigation of the process of dissemination of beans out of their domestication centers is on *P. vulgaris*, with considerably less knowledge, if any, regarding the other *Phaseolus* crop species.

For common bean, a very complex scenario was highlighted by numerous studies, which includes: (i) several introductions from the New World, in combination with exchanges between continents, and among several countries within continents; (ii) new agro-ecological conditions experienced by this crop, implying new opportunities for both natural and human-mediated selection to act; and (iii) loss of spatial isolation characteristic of the Americas, which allowed hybridization and introgression between the Andean and Mesoamerican gene pools, and as a consequence, the occurrence of novel genotypes and phenotypes that transgressed the parental phenotypes for important agronomic and adaptive traits, such as, e.g., nutritional quality and resistance to biotic and abiotic stresses (Angioi et al., 2010; Blair et al., 2010; Gioia et al., 2013a).

Patterns of Diversity of Beans Out of American Centers of Domestication

High levels of genetic diversity have been reported for common bean populations cultivated worldwide, and several continents and countries have been proposed as secondary centers of diversification for this species. These have included: the Iberian Peninsula (Santalla et al., 2002), the whole of Europe (Angioi et al., 2010, 2011; Gioia et al., 2013a), Brazil (Burle et al., 2010), central-eastern and southern Africa (Martin and Adams, 1987a,b; Asfaw et al., 2009; Blair et al., 2010), and China (Zhang et al., 2008).

In South America, a particular situation has emerged in Brazil. Although Brazil is closer to the Andes than to Mesoamerica, unexpectedly, it is the Mesoamerican *P. vulgaris* that is more prevalent (Burle et al., 2010). Multiple introductions of Mesoamerican germplasm in periods antecedent or successive to the discovery of the Americas might explain this pattern (Gepts et al., 1988).

In Africa, overall, the two gene pools are approximately equal in frequency, albeit there are strong differences between countries (Angioi et al., 2011; Okii et al., 2014). Such differences have been explained by the existence of at least partially independent seed networks in different countries (Asfaw et al., 2009), and because of selection due to dissimilar ecological and economic conditions among countries (Wortmann et al., 1998; Asfaw et al., 2009). Differential resistance to soil-borne diseases like

Fusarium root rot, and different yield performances arising from the 'interference' of improved genotypes released by national breeding programs have also been considered to explain the uneven distribution of the two gene pools across these regions (Blair et al., 2010).

In China, a prevalence of the Mesoamerican *P. vulgaris* has been observed (Zhang et al., 2008), although this was attributed mainly to founder effects (Zhang et al., 2008). The Himalayan region, as also the entire Indian subcontinent, shows high genetic diversity (Sofi et al., 2014; Rana et al., 2015).

Little is known about the dissemination of the common bean in India, albeit trading in the 16th century via the Red Sea and the Arabian Sea, and through the Hindustan Silk route, probably had a determinant role in the dissemination of this crop. It is also possible that the sea route discovered by the Portuguese explorer, Vasco da Gama, had a role in this dissemination. The genetic diversity in India also includes the combination of both the Mesoamerican and the Andean gene pools (Rana et al., 2015). Moreover, adaptation to micro-geographic conditions has been suggested for these landraces. Indeed, the analysis of >4000 landraces allowed the identification of several diverse clusters, irrespective of the place of collection, which also indicates a strong role for gene flow.

The Mesoamerican and Andean gene pools were both introduced into Europe. Studies carried out using several different marker types have shown that the Andean *P. vulgaris* predominates over Mesoamerican *P. vulgaris* (Gepts and Bliss, 1988; Lioi, 1989; Santalla et al., 2002; Logozzo et al., 2007; Angioi et al., 2010). The Andean type is largely predominant for the Iberian peninsula, Italy and central-northern Europe, where it is also prevalent on a local scale (Sicard et al., 2005; Angioi et al., 2009b). In the eastern part of Europe, the frequency of the Mesoamerican type tends to increase but it is always lower than the Andean (Papa et al., 2006). In their study of the expansion of the common bean *P. vulgaris* in Europe, Angioi et al. (2010) concluded that the intensity of the cytoplasmic bottleneck that resulted from this introduction into Europe was very low or absent (i.e., a loss of cpSSR diversity of ~2%).

Regarding the other *Phaseolus* crop species, there is little in the literature that has focused on the investigation of genetic diversity of *P. coccineus*, the allogamous sister species of *P. vulgaris*, in Europe. Some studies were conducted on small (Nowosielski et al., 2002; Sicard et al., 2005; Acampora et al., 2007; Boczkowska et al., 2012) and ample (Spataro et al., 2011; Rodriguez et al., 2013) spatial scales. The introduction of *P. coccineus* into Europe was probably contemporary with that of the common bean *P. vulgaris* (Westphal, 1974). Among the Mediterranean countries, *P. coccineus* is more widespread in Spain and Italy, while in northern Europe, it occurs more often in the UK and The Netherlands, where *P. coccineus* has often substituted for *P. vulgaris* (Santalla et al., 2004) as it is more adapted to cold temperatures and cool summers than *P. vulgaris* (Delgado Salinas, 1988; Rodino et al., 2007).

However, overall, *P. coccineus* is characterized by a narrower adaptability to environmental conditions than *P. vulgaris*. As previously mentioned, *P. coccineus* and *P. vulgaris* are cross-fertile when *P. vulgaris* is the maternal parent, and this might

have allowed hybridisation between these two species in Europe, where they often coexist in close sympatry in the same field. However, no evidence of introgression between common bean and runner bean has been found (Sicard et al., 2005). Data on nuclear and chloroplast variability of *P. coccineus* indicate that in Europe it has at least two main genetic groups (Spataro et al., 2011; Rodriguez et al., 2013). Of particular interest, there is a highly significant association between latitude and phenology for *P. coccineus* (Rodriguez et al., 2013). This relationship still holds when the effects of population structure for cpSSRs and nuSSRs is factored out. Therefore, this correlation is not just a consequence of the uneven geographic distribution of the two *P. coccineus* gene pools across Europe. It was then suggested that selection (probably for photoperiod sensitivity, and/or for low temperature), rather than migration and gene flow, has also had a role in shaping the population structure of *P. coccineus* in Europe (Rodriguez et al., 2013).

A comparison between Spanish and Mexican accessions of *P. vulgaris* and *P. coccineus* suggested that *P. coccineus* has maintained a high level of diversity since its introduction into Europe (Alvarez et al., 1998). A more recent study analyzed a worldwide collection of *P. coccineus*, and this analysis indicated that limited diversity of the runner bean *P. coccineus* appears to have been introduced into Europe, and that for nuclear markers, the European landraces show a reduction in diversity of 33% compared to that of the Mesoamerican landraces (Spataro et al., 2011). More recently, the use of chloroplast markers indicated a moderate-to-strong cytoplasmic bottleneck that followed the expansion of *P. coccineus* into Europe, with a reduction of 13% in chloroplast diversity (Rodriguez et al., 2013). As these markers are the same in number and type as those used by Angioi et al. (2010) to estimate the bottleneck of *P. vulgaris* following its expansion into Europe (2%), it can be concluded that the loss of diversity in *P. coccineus* appears to be stronger than in *P. vulgaris*.

Both nuclear and chloroplast analyses have shown that Mesoamerican and European *P. coccineus* accessions belong to distinct gene pools (Spataro et al., 2011; Rodriguez et al., 2013). It can be hypothesized that the differentiation of the European gene pool was due to adaptation to the new environment, and to genetic drift and a lack of introgression from wild forms.

The introduction of *P. lunatus* in Europe was very limited, with very few examples (Doria et al., 2012).

A Role for Adaptive Introgression for the Evolution of European Beans?

Using isozyme markers, Santalla et al. (2002) estimated a high percentage of hybrids in their common bean landrace collection from the Iberian peninsula (25%). Then later, through the integration of cytoplasmic and nuclear analyses, Angioi et al. (2010) reported that about 44% of their wide collection of landraces that spanned almost all of the European countries appeared to be derived from at least one hybridisation event. In addition to the molecular results, Angioi et al. (2010) showed that seed size and coat traits vary with the level of introgression between the two gene pools. More recently, Gioia et al. (2013a) used population assignment techniques to also reveal extensive

hybridisation between the two gene pools in Europe, with a frequency of hybridisation that was almost fourfold greater in Europe (40.2%) than in the Americas (12.3%), which confirmed the findings of Angioi et al. (2010). This can be explained by the geographic isolation between the gene pools in the American centers of origin, and that following the introduction into Europe, genotypes from different gene pools often coexisted on very small spatial scales (i.e., in small cultivated areas), and thus had the chance to hybridize. Estimations of hybridisation in other parts of the world have always been <10%, and so much less than in Europe (Zhang et al., 2008; Asfaw et al., 2009; Blair et al., 2010; Burle et al., 2010). Taken all together, this evidence supports the hypothesis that the whole of Europe can be regarded as a secondary center of diversity for the common bean (Angioi et al., 2010), as also suggested by the work of Santalla et al. (2002) that was limited to the Iberian peninsula.

After its introduction into Europe, *P. vulgaris* was exposed to new and ample agro-ecological conditions. Thus, it is likely that any releases that were not adapted to the new conditions were initially purged by selection. Among the natural factors, biotic and abiotic stress were probably determinants in shaping the genetic structure of the European bean landraces. It is possible that their adaptation to long days, their cold tolerance, and their resistance to pests and diseases were crucial, and this would probably have led to a reduction in the diversity that was initially present in the founding populations. Additionally, the selection operated by farmers for seed color and size, and culinary, organoleptic and nutritional quality might also have had strong impact on the evolution of the European bean, as witnessed by the myriad of local bean populations with particular characteristics and specific names (Angioi et al., 2009b; Lioi and Piergiovanni, 2013; De Ron et al., 2016). Moreover, the documented scenario of extensive Mesoamerican × Andean gene pools through their hybridisation in Europe (Logozzo et al., 2007; Angioi et al., 2010; Gioia et al., 2013a) suggests that introgressive hybridisation might have been the fundamental ‘evolutionary stimulus’ (Anderson and Stebbins, 1954; Lewontin and Birch, 1966) that propelled and boosted bean evolution in Europe. Indeed, hybridisation can produce new genotypic and phenotypic combinations that do not occur in either of the parental taxa, and upon which selection might act.

Recently, Bitocchi et al. (2015) reported that European flint landraces grown *in situ* show adaptive introgression from modern maize. A key result of their study was that adaptation followed by hybridisation has been very rapid, with landraces capturing and increasing the frequency of favorable alleles over very short times (e.g., 50 years) (Bitocchi et al., 2015). This allows the hypothesis that this evolutionary mechanism might also have operated for the European bean landraces, which have a history that is some 10-fold longer. This is of interest not only for studies in evolutionary genetics, but also for plant breeders. Indeed, studies focused on hybridisation have shown the potential for the identification of functionally important regions of the genome (Arnold and Martin, 2010).

The Possibility of Unraveling the Architecture of Adaptation of *Phaseolus vulgaris* in Europe

The potentiality of the studies at continental scale on the architecture of adaptation in common bean is confirmed by the possibility to apply a signature of selection approach using European landraces. Indeed, using methods for the identification of outlier loci for selection, Santalla et al. (2010) provided evidence that selective forces have had significant roles, particularly for seed size, growth habits, pest resistance and flowering time. This is intriguing, as selection for flowering time was probably a key element in the history of the bean also in its center of origin.

The BEAN_ADAPT project² is funded through the 2nd ERA-CAPS call, ERA-NET for Coordinating Action in Plant Sciences. The main aim of this project is to dissect out the genetic basis and phenotypic consequences of the adaptation to new environments of *P. vulgaris* and its sister species *P. coccineus*, through the study of their introduction from their respective centers of domestication in the Americas and their expansion through Europe as a recent and historically well-defined event of rapid adaptation. BEAN_ADAPT thus plans to characterize a large collection (11,500 accessions of each species) from three major genebanks by genotyping-by-sequencing. This will define the population structure and obtain subsets of genotypes for phenotyping (e.g., field, growth chamber) and for deeper genomic–transcriptomic–metabolomic characterisation. A multidisciplinary approach is planned (i.e., genomics, population/quantitative genetics, biochemistry, plant physiology) to unraveling the genetic bases of adaptation of these crops in new agro-ecological environments. The methods here rely on previous studies that have demonstrated the effectiveness of various analysis, such as the application of: (i) population genomics, to test for signatures of selection (e.g., Bellucci et al., 2014a); (ii) evolutionary metabolomics, which appear to be a very powerful approach to characterize molecular phenotypic changes due to domestication, and to identify traits under selection in wheat (Beleggia et al., 2016); (iii) association mapping studies of complex traits, such as flowering time (Raggi et al., 2014); (iv) analysis of signatures of selection by searching for ‘unusually high’ correlations between SNPs and environmental variables at the continental scale, as successfully applied by Rodriguez et al. (2016) for wild common bean from Mexico.

CONCLUSION

This review presents a comprehensive overview of the current knowledge about the evolutionary history of the *Phaseolus* crop species. This takes us from the origins and evolution of their wild forms, with their co-evolution and interactions with humans and diverse environments during and after the domestication process, to their colonization of new environments

out of their centers of origin and domestication. The picture that has unraveled shows that the specific and almost unique features of the *Phaseolus* genus make it a very powerful model to address important evolutionary issues. In particular, among crops, *Phaseolus* represents a unique example of multiple parallel domestications for five closely related species, with two of these (i.e., *P. vulgaris* and *P. lunatus*) each domesticated independently in Mesoamerica and the Andes. This has resulted in at least a total of seven independent domestication events, involving species that have diverged relatively recently (~4 Ma; Delgado-Salinas et al., 2006) and that show similar genomic structures, making inter-specific comparisons feasible. This represents a ‘domestication experiment’ with full factorial design for three factors, as species, areas and wild/domesticated status, which can provide a deep understanding of the genomic architecture of domestication. Moreover, the study of different domestication events provides valuable replicates for the understanding of convergent evolution (i.e., different species or populations that evolved similar phenotypes) and its genomic determinants and effects. The possibility of considering these multiple parallel domestications as different replicates of the same experiment is also supported by the recent findings of Vlasova et al. (2016), who reported that both lineages of *P. vulgaris* and potentially all of the *Phaseolus* species share the same patterns of gene duplication that predate the divergence between the Mesoamerican and Andean gene pools.

Along with domestication, the complex patterns of dissemination of *Phaseolus* crops out of their centers of domestication represents a further key strength to unravel the genetic basis of plant adaptation. This is exactly what the BEAN_ADAPT project is searching for: to dissect out the genetic basis and phenotypic consequences of the adaptation to new environments of the common bean and its sister species, the runner bean, through the study of their introduction and expansion through Europe, as a recent and historically well-defined event of rapid adaptation.

This approach can provide a model for future major environmental and socio-economic changes, such as increases in temperature, variability of rainfall, and new consumer preferences, which will be fruitful for both evolutionary biologists and plant breeders. For the evolutionary biologists, it will be of particular interest to compare the results obtained across different species and populations, to look for patterns of convergent evolution, at either the phenotypic or molecular level. Discovering genes and genetic mechanisms that contribute to phenotypic adaptation associated with environmental conditions and their mapping along the reference genome will provide a useful genetic tool for geneticists and breeders for the design of novel varieties.

AUTHOR CONTRIBUTIONS

RP and EBi designed and wrote the manuscript. EBe, DS, LN, EBi, and RP contributed to the drafting of the manuscript especially for the sections on the origin and domestication of *Phaseolus* spp.,

²<http://www.beanadapt.org/>

while DR, MM, MR, GA and TG contributed to the drafting of the manuscript especially for the sections on dissemination out of the domestication centres of *Phaseolus* ssp.

ACKNOWLEDGMENTS

This work was supported by grants from the ERA-NET for Coordinating Action in Plant Sciences-2nd ERA-CAPS call, BEAN_ADAPT project, the Italian Government (MIUR; Grant number RBFR13IDFM_001, FIRB Project 2013) and

the Università Politecnica delle Marche (years 2015–2016). All of the appropriate permissions for reproduction of previously published Figures have been obtained from the respective copyright holders.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00722/full#supplementary-material>

REFERENCES

- Acampora, A., Ciaffi, M., De Pace, C., Paolacci, A. R., and Tanzarella, O. A. (2007). Pattern of variation for seed size traits and molecular markers in Italian germplasm of *Phaseolus coccineus* L. *Euphytica* 157, 69–82. doi: 10.1007/s10681-007-9397-3
- Allard, R. W. (1960). "Patterns of evolution in cultivated species," in *Principles of Plant Breeding*, ed. R. W. Allard (New York, NY: Wiley), 7–18.
- Alvarez, M. T., Sáenz de Miera, L. E., and Pérez de la Vega, M. (1998). Genetic variation in common and runner bean of the Northern Meseta in Spain. *Genet. Res. Crop Evol.* 45, 243–251. doi: 10.1023/A:1008604127933
- Alva-Valdivia, L. M., Goguitchaichvili, A., Ferrari, L., Rosas-Elguera, J., Urrutia-Fucugauchi, J., and Zamorano-Orozco, J. J. (2000). Paleomagnetic data from the trans-mexican volcanic belt: implications for tectonics and volcanic stratigraphy. *Earth Planets Space* 52, 467–478. doi: 10.1186/BF03351651
- Anderson, E., and Stebbins, G. L. Jr. (1954). Hybridization as an evolutionary stimulus. *Evolution* 8, 378–388. doi: 10.2307/2405784
- Andueza-Noh, R. H., Martínez-Castillo, J., and Chacón-Sánchez, M. I. (2015). Domestication of small-seeded lima bean (*Phaseolus lunatus* L.) landraces in Mesoamerica: evidence from microsatellite markers. *Genetica* 143, 657–669. doi: 10.1007/s10709-015-9863-0
- Andueza-Noh, R. H., Serrano-Serrano, M. L., Chacón, M. I., Sánchez del-Pino, I., Camacho-Pérez, L., Coello-Coello, J., et al. (2013). Multiple domestications of the Mesoamerican gene pool of Lima bean (*Phaseolus lunatus* L.): evidence from chloroplast DNA sequences. *Genet. Res. Crop Evol.* 60, 1069–1086. doi: 10.1007/s10722-012-9904-9
- Angioi, S. A., Desiderio, F., Rau, D., Bitocchi, E., Attene, G., and Papa, R. (2009a). Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biol.* 11, 598–612. doi: 10.1111/j.1438-8677.2008.00143.x
- Angioi, S. A., Rau, D., Rodríguez, M., Logozzo, G., Desiderio, F., Papa, R., et al. (2009b). Nuclear and chloroplast microsatellite diversity in *Phaseolus vulgaris* L. from Sardinia (Italy). *Mol. Breed.* 23, 413–429. doi: 10.1007/s11032-008-9245-8
- Angioi, S. A., Rau, D., Attene, G., Nanni, L., Bellucci, E., Logozzo, G., et al. (2010). Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. *Theor. Appl. Genet.* 121, 829–843. doi: 10.1007/s00122-010-1353-2
- Angioi, S. A., Rau, D., Nanni, L., Bellucci, E., Papa, R., and Attene, G. (2011). The genetic make-up of the European landraces of the common bean. *Plant Genet. Res.* 9, 197–201. doi: 10.1017/S1479262111000190
- Arnold, M. L., and Martin, N. H. (2010). Hybrid fitness across time and habitats. *Trends Ecol. Evol.* 25, 530–536. doi: 10.1016/j.tree.2010.06.005
- Asfaw, A., Blair, M. W., and Almekinders, C. (2009). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor. Appl. Genet.* 120, 1–12. doi: 10.1007/s00122-009-1154-7
- Baudoin, J. P. (1988). "Genetic resources, domestication and evolution of lima bean, *Phaseolus lunatus*," in *Genetic Resources of Phaseolus Beans*, ed. P. Gepts (Berlin: Springer), 393–407.
- Beebe, S., Rengifo, J., Gaitan, E., Duque, M. C., and Tohme, J. (2001). Diversity and origin of Andean landraces of common bean. *Crop Sci.* 41, 854–862. doi: 10.2135/cropsci2001.413854x
- Beleggia, R., Rau, D., Laidò, G., Platani, C., Nigro, F., Fragasso, M., et al. (2016). Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Mol. Biol. Evol.* 33, 1740–1753. doi: 10.1093/molbev/msw050
- Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S., et al. (2014a). Decreased nucleotide and expression diversity and modified coexpression patterns characterize domestication in the common bean. *Plant Cell* 26, 1901–1912.
- Bellucci, E., Bitocchi, E., Rau, D., Rodríguez, M., Biagetti, E., Giardini, A., et al. (2014b). "Genomics of origin, domestication and evolution of *Phaseolus vulgaris*," in *Genomics of Plant Genetic Resources*, eds R. Tuberosa, A. Graner, and E. Frison (Berlin: Springer), 483–507. doi: 10.1007/978-94-007-7572-5_20
- Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodríguez, M., Biagetti, E., et al. (2013). Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol.* 197, 300–313. doi: 10.1111/j.1469-8137.2012.04377.x
- Bitocchi, E., Bellucci, E., Rau, D., Albertini, E., Veronesi, F., Attene, G., et al. (2015). European flint landraces grown in situ reveal adaptive introgression from modern maize. *PLoS ONE* 10:e0121381. doi: 10.1371/journal.pone.0121381
- Bitocchi, E., Nanni, L., Bellucci, E., Rossi, M., Giardini, A., Spagnoletti Zeuli, P. S., et al. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 109, E788–E796. doi: 10.1073/pnas.1108973109
- Bitocchi, E., Rau, D., Benazzo, A., Bellucci, E., Goretti, D., Biagetti, E., et al. (2016a). High level of nonsynonymous changes in common bean suggests that selection under domestication increased functional diversity at target traits. *Front. Plant Sci.* 7:2005. doi: 10.3389/fpls.2016.02005
- Bitocchi, E., Rau, D., Rodríguez, M., and Murgia, M. L. (2016b). "Crop improvement of *Phaseolus* spp. through interspecific and intraspecific hybridization," in *Polyploidy and Hybridization for Crop Improvement*, ed. A. Mason (Boca Raton, FL: CRC Press), 218–280. doi: 10.1201/9781315369259-11
- Blair, M. W., González, L. F., Kimani, M., and Butare, L. (2010). Genetic diversity, inter-gene pool introgression and nutritional quality of common beans (*Phaseolus vulgaris* L.) from Central Africa. *Theor. Appl. Genet.* 121, 237–248. doi: 10.1007/s00122-010-1305-x
- Blair, M. W., Muñoz, L. C., and Debouck, D. G. (2002). Tepary beans (*P. acutifolius*): molecular analysis of a forgotten genetic resource for dry land agriculture. *Grain Legumes* 36, 25–26.
- Blair, M. W., Pantoja, W., and Muñoz, L. C. (2012). First use of microsatellite markers in a large collection of cultivated and wild accessions of tepary bean (*Phaseolus acutifolius* A. Gray). *Theor. Appl. Genet.* 125, 1137–1147. doi: 10.1007/s00122-012-1900-0
- Boczkowska, M., Bulinska-Radomska, Z., and Nowosielski, J. (2012). AFLP analysis of genetic diversity in five accessions of Polish runner bean (*Phaseolus coccineus* L.). *Genet. Res. Crop Evol.* 59, 473–478. doi: 10.1007/s10722-012-9798-6
- Bonifácio, E. M., Fonsêca, A., Almeida, C., Santos, K. G. B., and Pedrosa-Harand, A. (2012). Comparative cytogenetic mapping between the lima bean (*Phaseolus lunatus* L.) and the common bean (*P. vulgaris* L.). *Theor. Appl. Genet.* 124, 1513–1520. doi: 10.1007/s00122-012-1806-x
- Bradburd, G. S., Ralph, P. L., and Coop, G. M. (2013). Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* 67, 3258–3273. doi: 10.1111/evo.12193
- Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.): model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611

- Brown, C. H., Clement, C. R., Epps, P., Luedeling, E., and Wichmann, S. (2014). The paleobiolinguistics of the common bean (*Phaseolus vulgaris* L.). *Ethnobiol. Lett.* 5, 104–115. doi: 10.124237/eb1.5.2014.203
- Burle, M. L., Fonseca, J. R., Kami, J. A., and Gepts, P. (2010). Microsatellite diversity and genetic structure among common bean (*Phaseolus vulgaris* L.) landraces in Brazil, a secondary center of diversity. *Theor. Appl. Genet.* 121, 801–813. doi: 10.1007/s00122-010-1350-5
- Caicedo, A. L., Gaitán, E., Duque, M. C., Toro Chica, O., Debouck, D. G., and Thome, J. (1999). AFLP fingerprinting of *Phaseolus lunatus* L. and related wild species from South America. *Crop Sci.* 39, 1497–1507. doi: 10.2135/cropsci1999.3951497x
- Cavalli-Sforza, L. (1966). Population structure and human evolution. *Proc. R. Soc. Lond. B Biol. Sci.* 164, 362–379. doi: 10.1098/rspb.1966.0038
- Chacón, S. M. I., Pickersgill, B., and Debouck, D. G. (2005). Domestication patterns in common bean (*Phaseolus vulgaris* L.) and the origin of the Mesoamerican and Andean cultivated races. *Theor. Appl. Genet.* 110, 432–444. doi: 10.1007/s00122-004-1842-2
- Choi, J. Y., Platts, A. E., Fuller, D. Q., Hsing, Y. I., Wing, R. A., and Purugganan, M. D. (2017). The rice paradox: multiple origins but single domestication in Asian rice. *Mol. Biol. Evol.* 34, 969–979. doi: 10.1093/molbev/msx049
- Darwin, C. (1859). *On the Origins of Species by Means of Natural Selection*. London: John Murray.
- de Alencar Figueiredo, L. F., Calatayud, C., Dupuits, C., Billot, C., Rami, J. F., Brunel, D., et al. (2008). Phylogeographic evidence of crop neodiversity in sorghum. *Genetics* 179, 997–1008. doi: 10.1534/genetics.108.087312
- De Ron, A. M., González, A. M., Rodiño, A. P., Santalla, M., Godoy, L., and Papa, R. (2016). History of the common bean crop: its evolution beyond its areas of origin and domestication. *ARBOR* 192, a317. doi: 10.3989/arbor.2016.779n3007
- Debouck, D. G., Toro, O., Paredes, O. M., Johnson, W. C., and Gepts, P. (1993). Genetic diversity and ecological distribution of *Phaseolus vulgaris* in northwestern South America. *Econ. Bot.* 47, 408–423. doi: 10.1007/BF02907356
- Delgado Salinas, A. (1988). “Variation, taxonomy, domestication and germplasm potentialities in *Phaseolus coccineus*,” in *Genetic Resources of Phaseolus Beans*, ed. P. Gepts (Berlin: Springer), 441–463.
- Delgado-Salinas, A., Bibler, R., and Lavin, M. (2006). Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Syst. Bot.* 31, 779–791. doi: 10.1600/036364406779695960
- Delgado-Salinas, A., Turley, T., Richman, A., and Lavin, M. (1999). Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). *Syst. Bot.* 24, 438–460. doi: 10.2307/2419699
- Desiderio, F., Bitocchi, E., Bellucci, E., Rau, D., Rodriguez, M., Attene, G., et al. (2013). Chloroplast microsatellite diversity in *Phaseolus vulgaris*. *Front. Plant Sci.* 3:312. doi: 10.3389/fpls.2012.00312
- Doria, E., Campion, B., Sparvoli, F., Tava, A., and Nielsen, E. (2012). Anti-nutrient components and metabolites with health implications in seeds of 10 common bean (*Phaseolus vulgaris* L. and *Phaseolus lunatus* L.) landraces cultivated in southern Italy. *J. Food Compos. Anal.* 26, 72–80. doi: 10.1016/j.jfca.2012.03.005
- Escalante, A., Coello, G., Eguarte, L. E., and Piñero, D. (1994). Genetic structure and mating systems in wild and cultivated populations of *Phaseolus coccineus* and *P. vulgaris* (Fabaceae). *Am. J. Bot.* 81, 1096–1103. doi: 10.2307/2445471
- Fofana, B., Baudoin, J. P., Vekemans, X., Debouck, D. G., and du Jardin, P. (1999). Molecular evidence for an Andean origin and a secondary gene pool for the Lima bean (*Phaseolus lunatus*) using chloroplast DNA. *Theor. Appl. Genet.* 98, 202–212. doi: 10.1007/s001220051059
- Fofana, B., du Jardin, P., and Baudoin, J. P. (2001). Genetic diversity in the Lima bean (*Phaseolus lunatus* L.) as revealed by chloroplast DNA (cpDNA) variations. *Genet. Res. Crop Evol.* 48, 437–445. doi: 10.1023/A:1012016030318
- Fonseca, A., and Pedrosa-Harand, A. (2013). Karyotype stability in the genus *Phaseolus* evidenced by the comparative mapping of the wild species *Phaseolus microcarpus*. *Genome* 56, 335–343. doi: 10.1139/gen-2013-2025
- Freeman, G. F. (1912). *Southwestern Beans and Teparies*, Vol. 68. Tucson, AZ: University of Arizona.
- Freeman, G. F. (1913). The tepary, a new cultivated legume from the Southwest. *Bot. Gaz.* 56, 395–417. doi: 10.1086/331183
- Freytag, G. F., and Debouck, D. G. (2002). *Taxonomy, Distribution, and Ecology of the Genus Phaseolus (Leguminosae-Papilionoideae) in North America, Mexico and Central America*. Fort Worth, TX: Botanical Research Institute Of Texas.
- Garvin, D. F., and Weeden, N. F. (1994). Isozyme evidence supporting a single geographic origin for domesticated tepary bean. *Crop Sci.* 34, 1390–1395. doi: 10.2135/cropsci1994.0011183X003400050045x
- Gepts, P. (1988). “A Middle American and an Andean common bean gene pool,” in *Genetic Resources of Phaseolus Beans*, ed. P. Gepts (Dordrecht: Springer), 375–390. doi: 10.1007/978-94-009-2786-5_15
- Gepts, P., and Bliss, F. A. (1988). Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. II Europe and Africa. *Econ. Bot.* 42, 86–104. doi: 10.1007/BF02859038
- Gepts, P., Kmiecik, K., Pereira, P., and Bliss, F. A. (1988). Dissemination pathways of common bean (*Phaseolus vulgaris*, Fabaceae) deduced from phaseolin electrophoretic variability. I. The Americas. *Econ. Bot.* 42, 73–85. doi: 10.1007/BF02859036
- Gepts, P., Papa, R., Coulibaly, S., Mejía, A. G., and Pasquet, R. (1999). “Wild legume diversity and domestication - insights from molecular methods,” in *Wild Legumes Proceedings 7th MAFF International Workshop on Genetic Resources*, ed. D. Vaughan (Tsukuba: National Institute of Agrobiological Resources), 19–31.
- Gioia, T., Logozzo, G., Attene, G., Bellucci, E., Benedettelli, S., Negri, V., et al. (2013a). Evidence for introduction bottleneck and extensive inter-gene pool (Mesoamerica x Andes) hybridization in the European common bean (*Phaseolus vulgaris* L.) germplasm. *PLoS ONE* 8:e75974. doi: 10.1371/journal.pone.0075974
- Gioia, T., Logozzo, G., Kami, J., Spagnoletti Zeuli, P., and Gepts, P. (2013b). Identification and characterization of a homologue to the *Arabidopsis INDEHISCENT* gene in common bean. *J. Hered.* 104, 273–286. doi: 10.1093/jhered/ess102
- Glémin, S., and Bataillon, T. (2009). A comparative view of the evolution of grasses under domestication. *New Phytol.* 183, 273–290. doi: 10.1111/j.1469-8137.2009.02884.x
- Gregory-Wodzicki, K. M. (2000). Uplift history of the central and northern Andes: a review. *Geol. Soc. Am. Bull.* 112, 1091–1105. doi: 10.1130/0016-7606
- Guillot, G., Vitalis, R., le Rouzic, A., and Gautier, M. (2014). Detecting correlation between allele frequencies and environmental variables as a signature of selection. A fast computational approach for genome-wide studies. *Spat. Stat.* 8, 145–155. doi: 10.1016/j.spa.2013.08.001
- Gujaria-Verma, N., Ramsay, L., Sharpe, A. G., Sanderson, L. A., Debouck, D. G., Taran, B., et al. (2016). Gene-based SNP discovery in tepary bean (*Phaseolus acutifolius*) and common bean (*P. vulgaris*) for diversity analysis and comparative mapping. *BMC Genomics* 17:239. doi: 10.1186/s12864-016-2499-3
- Gutiérrez-Salgado, A., Gepts, P., and Debouck, D. G. (1995). Evidence for two gene pools of the Lima beans, *Phaseolus lunatus* L., in the Americas. *Genet. Res. Crop Evol.* 42, 15–28. doi: 10.1007/BF02310680
- Harlan, J. R. (1975). *Crops and Man*. Madison, WI: American Society of Agronomy.
- Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., et al. (2005). Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiol.* 137, 1420–1434. doi: 10.1104/pp.104.057018
- Hecht, V., Knowles, C. L., Vander Schoor, J. K., Liew, L. C., Jones, S. E., Lambert, M. J., et al. (2007). Pea LATE BLOOMER1 is a GIGANTEA ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant Physiol.* 144, 648–661. doi: 10.1104/pp.107.096818
- Heiser, C. B. (1965). Cultivated plants and cultural diffusion in nuclear America. *Am. Anthropol.* 67, 930–949. doi: 10.1525/aa.1965.67.4.02a00040
- Hucl, P., and Scoles, G. J. (1985). Interspecific hybridization in the common bean: a review. *HortScience* 20, 352–357.
- Hufford, M. B., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J. M., Cartwright, R. A., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309
- Kami, J., Velásquez, V. B., Debouck, D. G., and Gepts, P. (1995). Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1101–1104. doi: 10.1073/pnas.92.4.1101
- Kaplan, L., and Lynch, T. (1999). *Phaseolus* (Fabaceae) in archaeology: AMS radiocarbon dates and their significance for pre-Columbian agriculture. *Econ. Bot.* 53, 261–272. doi: 10.1007/BF02866636
- Kim, Y., Yeom, M., Kim, H., Lim, J., Koo, H. J., Hwang, D., et al. (2012). GIGANTEA and EARLY FLOWERING 4 in *Arabidopsis* exhibit differential

- phase-specific genetic influences over a diurnal cycle. *Mol. Plant* 5, 152–161. doi: 10.1093/mp/sss005
- Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x
- Kraft, K. H., Brown, C. H., Nabhan, G. P., Luedeling, E., Ruiz, J. D. J. L., Coppens d'Eeckenbrugge, G., et al. (2014). Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6165–6170. doi: 10.1073/pnas.1308933111
- Kwak, M., and Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* 118, 979–992. doi: 10.1007/s00122-008-0955-4
- Kwak, M., Kami, J., and Gepts, P. (2009). The putative Mesoamerican domestication center of *Phaseolus vulgaris* is located in the Lerma-Santiago Basin of Mexico. *Crop Sci.* 49, 554–563. doi: 10.2135/cropsci2008.07.0421
- Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Lester, R. N. (1989). Evolution under domestication involving disturbance of genic balance. *Euphytica* 44, 125–132. doi: 10.1007/BF00022606
- Lewontin, R. C., and Birch, L. C. (1966). Hybridization as a source of variation for adaptation to new environments. *Evolution* 20, 315–336. doi: 10.2307/2406633
- Lewontin, R. C., and Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphism. *Genetics* 74, 175–195.
- Liljgren, S., Ditta, G., Eshed, Y., Savidge, B., Bowmanand, J., and Yanofsky, M. (2000). Control of fruit dehiscence in *Arabidopsis* by the SHATTERPROOF MADS-box genes. *Nature* 404, 766–769. doi: 10.1038/35008089
- Lin, Z., Li, X., Shannon, L. M., Yeh, C. T., Wang, M. L., Bai, G., et al. (2012). Parallel domestication of the *Shattering1* genes in cereals. *Nat. Genet.* 44, 720–724. doi: 10.1038/ng.2281
- Lioi, L. (1989). Geographical variation of phaseolin patterns in an old world collection of *Phaseolus vulgaris*. *Seed Sci. Technol.* 17, 317–324.
- Lioi, L., Lotti, C., and Galasso, I. (1998). Isozyme diversity, RFLP of the rDNA and phylogenetic affinities among cultivated lima beans, *Phaseolus lunatus* (Fabaceae). *Plant Syst. Evol.* 213, 153–164. doi: 10.1007/BF00985196
- Lioi, L., and Piergiovanni, A. R. (2013). “European common bean,” in *Genetic and Genomic Resources of Grain Legume Improvement*, eds M. Singh, H. D. Upadhyaya, and I. S. Bisht (London: Elsevier Inc.), 11–40. doi: 10.1016/B978-0-12-397935-3.00002-5
- Lippold, S., Knapp, M., Kuznetsova, T., Leonard, J. A., Benecke, N., Ludwig, A., et al. (2011). Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nat. Commun.* 2, 450. doi: 10.1038/ncomms1447
- Logozzo, G., Donnoli, R., Macaluso, L., Papa, R., Knüpfner, H., and Spagnoletti Zeuli, P. S. (2007). Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genet. Res. Crop Evol.* 54, 1763–1779. doi: 10.1007/s10722-006-9185-2
- Londo, J. P., Chiang, Y. C., Hung, K. H., Chiang, T. Y., and Schaal, B. A. (2006). Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 9578–9583. doi: 10.1073/pnas.0603152103
- Lu, J., Tang, T., Tang, H., Huang, J., Shi, S., and Wu, C. I. (2006). The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* 22, 126–131. doi: 10.1016/j.tig.2006.01.004
- Mamidi, S., Rossi, M., Annam, D., Moghaddam, S., Lee, R., Papa, R., et al. (2011). Investigation of the domestication of common bean (*Phaseolus vulgaris*) using multilocus sequence data. *Funct. Plant Biol.* 38, 953–967. doi: 10.1071/FP11124
- Mamidi, S., Rossi, M., Moghaddam, S. M., Annam, D., Lee, R., Papa, R., et al. (2013). Demographic factors shaped diversity in the two gene pools of wild common bean *Phaseolus vulgaris* L. *Heredity* 110, 267–276. doi: 10.1038/hdy.2012.82
- Manshardt, R. M., and Waines, J. G. (1983). Isozyme variation and the origin of domesticated tepary beans (*Phaseolus acutifolius* Gray). *Annu. Rep. Bean Improv. Coop.* 26, 18–19.
- Maquet, A., Zoro Bi, I., Delvaux, M., Wathélet, B., and Baudoin, J. P. (1997). Genetic structure of a lima bean base collection using allozyme markers. *Theor. Appl. Genet.* 95, 980–991. doi: 10.1007/s001220050651
- Márquez, A., Verma, S. P., Anguita, F., Oyarzun, R., and Brandle, J. L. (1999). Tectonics and volcanism of sierra chichinautzin: extension at the front of the central trans-mexican volcanic belt. *J. Volcanol. Geotherm. Res.* 93, 125–150. doi: 10.1016/S0377-0273(99)00085-2
- Martin, G. B., and Adams, M. W. (1987a). Landraces of *Phaseolus vulgaris* (Fabaceae) in northern Malawi I. Regional variation. *Econ. Bot.* 41, 190–203. doi: 10.1007/BF02858965
- Martin, G. B., and Adams, M. W. (1987b). Landraces of *Phaseolus vulgaris* (Fabaceae) in northern Malawi II. Generation and maintenance of variability. *Econ. Bot.* 41, 204–215. doi: 10.1007/BF02858967
- Martínez-Castillo, J., Camacho-Pérez, L., Villanueva-Viramontes, S., Andueza-Noh, R. H., and Chacón-Sánchez, M. I. (2014). Genetic structure within the Mesoamerican gene pool of wild *Phaseolus lunatus* (Fabaceae) from Mexico as revealed by microsatellite markers: implications for conservation and the domestication of the species. *Am. J. Bot.* 101, 851–864. doi: 10.3732/ajb.1300412
- Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, G. J., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6080–6084. doi: 10.1073/pnas.052125199
- McCouch, S. (2004). Diversifying selection in plant breeding. *PLoS Biol.* 2:e347. doi: 10.1371/journal.pbio.0020347
- Mendel, G. (1866). Versuche über Pflanzenhybriden. *Verh. Naturf. Verein. Brünn* 4, 1–47.
- Messina, V. (2014). Nutritional and health benefits of dried beans. *Am. J. Clin. Nutr.* 100, 437S–442S. doi: 10.3945/ajcn.113.071472
- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Mizoguchi, T., Wright, L., Fujiwara, S., Cremer, F., Lee, K., Onouchi, H., et al. (2005). Distinct roles of GIGANTEA in promoting flowering and regulating circadian rhythms in *Arabidopsis*. *Plant Cell* 17, 2255–2270. doi: 10.1105/tpc.105.033464
- Molina, J., Sikora, M., Garud, N., Flowers, J. M., Rubinstein, S., Reynolds, A., et al. (2011). Molecular evidence for a single evolutionary origin of domesticated rice. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8351–8356. doi: 10.1073/pnas.1104686108
- Motta-Aldana, J., Serrano-Serrano, M. L., Torres, H. J., Villamizar, C. G., Debouck, D. G., and Chacón, M. I. (2010). Multiple origins of Lima bean landraces in the Americas: evidence from chloroplast and nuclear DNA polymorphisms. *Crop Sci.* 50, 1773–1787. doi: 10.2135/cropsci2009.12.0706
- Muñoz, L. C., Duque, M. C., Debouck, D. G., and Blair, M. W. (2006). Taxonomy of tepary bean and wild relatives as determined by amplified fragment length polymorphism (AFLP) markers. *Crop Sci.* 46, 1744–1754. doi: 10.2135/cropsci2005-12-0475
- Murgia, M. L., Attene, G., Rodriguez, M., Bitocchi, E., Bellucci, E., Fois, D., et al. (2017). A comprehensive phenotypic investigation of the ‘pod-shattering syndrome’ in common bean. *Front. Plant Sci.* 8:251. doi: 10.3389/fpls.2017.00251
- Nabhan, G. P., and Felger, R. S. (1978). Teparies in southwestern North America. *Econ. Bot.* 32, 3–19. doi: 10.1007/BF02906725
- Nanni, L., Bitocchi, E., Bellucci, E., Rossi, M., Rau, D., Attene, G., et al. (2011). Nucleotide diversity of a genomic sequence similar to SHATTERPROOF (PvSH1) in domesticated and wild common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 123, 1341–1357. doi: 10.1007/s00122-011-1671-z
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590.
- Nieto-Samaniego, A., Ferrari, L., Alaniz-Alvarez, S., Labarthe-Hernandez, G., and Rosas-Elguera, J. (1999). Variation of Cenozoic extension and volcanism across the southern Sierra Madre Occidental volcanic province, Mexico. *Geol. Soc. Am. Bull.* 111, 347–363. doi: 10.1130/0016-7606
- Nowosielski, J., Podyma, W., and Nowosielska, D. (2002). Molecular research on the genetic diversity of Polish varieties and landraces of *Phaseolus coccineus* L. and *Phaseolus vulgaris* L. using the RAPD and AFLP methods. *Cell. Mol. Biol. Lett.* 7, 753–762.
- Okii, D., Tukamuhawa, P., Kami, J., Namayanja, A., Paparu, P., Ugen, M., et al. (2014). The genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) germplasm in Uganda. *Afr. J. Biotechnol.* 13, 2935–2949. doi: 10.5897/AJB2014.13916

- Olson, M. V. (1999). When less is more: gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* 64, 18–23. doi: 10.1086/302219
- Osakabe, Y., Arinaga, N., Umezawa, T., Katsura, S., Nagamachi, K., Tanaka, H., et al. (2013). Osmotic stress responses and plant growth controlled by potassium transporters in *Arabidopsis*. *Plant Cell* 25, 609–624. doi: 10.1105/tpc.112.105700
- Papa, R., Acosta, J., Delgado-Salinas, A., and Gepts, P. (2005). A genome-wide analysis of differentiation between wild and domesticated *Phaseolus vulgaris* from Mesoamerica. *Theor. Appl. Genet.* 111, 1147–1158. doi: 10.1007/s00122-005-0045-9
- Papa, R., Bellucci, E., Rossi, M., Leonardi, S., Rau, D., Gepts, P., et al. (2007). Tagging the signatures of domestication in common bean (*Phaseolus vulgaris*) by means of pooled DNA samples. *Ann. Bot.* 100, 1039–1051. doi: 10.1093/aob/mcm151
- Papa, R., and Gepts, P. (2003). Asymmetry of gene flow and differential geographical structure of molecular diversity in wild and domesticated common bean (*Phaseolus vulgaris* L.) from mesoamerica. *Theor. Appl. Genet.* 106, 239–250. doi: 10.1007/s00122-002-1085-z
- Papa, R., Nanni, L., Sicard, D., Rau, D., and Attene, G. (2006). “The evolution of genetic diversity in *Phaseolus vulgaris* L.” in *New Approaches to the Origins, Evolution and Conservation of Crops. Darwin's Harvest*, eds T. J. Motley, N. Zerega, and H. Cross (New York, NY: Columbia University Press), 121–142.
- Plunket, P., and Uruñuela, G. (2012). Where east meets west: the formative in Mexico's central highlands. *J. Archaeol. Res.* 20, 1–51. doi: 10.1007/s10814-011-9051-4
- Purugganan, M. D., and Fuller, D. (2009). The nature of selection during plant domestication. *Nature* 457, 843–848. doi: 10.1038/nature07895
- Raggi, L., Tissi, C., Mazzucato, A., and Negri, V. (2014). Molecular polymorphism related to flowering trait variation in a *Phaseolus vulgaris* L. collection. *Plant Sci.* 215, 180–189. doi: 10.1016/j.plantsci.2013.11.001
- Ramírez-Villegas, J., Khoury, C., Jarvis, A., Debouck, D. G., and Guarino, L. (2010). A gap analysis methodology for collecting crop gene pools: a case study with *Phaseolus* beans. *PLoS ONE* 5:e13497. doi: 10.1371/journal.pone.0013497
- Rana, J. C., Sharma, T. R., Tyagi, R. K., Chahota, R. K., Gautam, N. K., Singh, M., et al. (2015). Characterisation of 4274 accessions of common bean (*Phaseolus vulgaris* L.) germplasm conserved in the Indian gene bank for phenological, morphological and agricultural traits. *Euphytica* 205, 441–457. doi: 10.1007/s10681-015-1406-3
- Rendón-Anaya, M., Montero-Vargas, J. M., Saburido-Álvarez, S., Vlasova, A., Capella-Gutiérrez, S., Ordaz-Ortiz, J. J., et al. (2017). Genomic history of the origin and domestication of common bean unveils its closest sister species. *Genome Biol.* 18, 60. doi: 10.1186/s13059-017-1190-6
- Rodino, A. P., Lema, E. M., Marlene, P. B., Santalla, M., and De Ron, A. M. (2007). Assessment of runner bean (*Phaseolus coccineus* L.) germplasm for tolerance to low temperature during early seedling growth. *Euphytica* 155, 63–70. doi: 10.1007/s10681-006-9301-6
- Rodríguez, M., Rau, D., Angioi, S. A., Bellucci, E., Bitocchi, E., Nanni, L., et al. (2013). European *Phaseolus coccineus* L. landraces: population structure and adaptation, as revealed by cpSSRs and phenotypic analyses. *PLoS ONE* 8:e57337. doi: 10.1371/journal.pone.0057337
- Rodríguez, M., Rau, D., Bitocchi, E., Bellucci, E., Biagetti, E., Carboni, A., et al. (2016). Landscape genetics, adaptive diversity, and population structure in *Phaseolus vulgaris*. *New Phytol.* 209, 1781–1794. doi: 10.1111/nph.13713
- Rossi, M., Bitocchi, E., Bellucci, E., Nanni, L., Rau, D., Attene, G., et al. (2009). Linkage disequilibrium and population structure in wild and domesticated populations of *Phaseolus vulgaris* L. *Evol. Appl.* 2, 504–522. doi: 10.1111/j.1752-4571.2009.00082.x
- Rubiales, D., and Mikic, A. (2015). Introduction: legumes in sustainable agriculture. *Crit. Rev. Plant Sci.* 34, 2–3. doi: 10.1080/07352689.2014.897896
- Salunkhe, D. K., and Kadam, S. S. (1998). *Handbook of Vegetable Science and Technology: Production, Compostion, Storage, and Processing*. Boca Raton, FL: CRC press.
- Santalla, M., De Ron, A. M., and De La Fuente, M. (2010). Integration of genome and phenotypic scanning gives evidence of genetic structure in Mesoamerican common bean (*Phaseolus vulgaris* L.) landraces from the southwest of Europe. *Theor. Appl. Genet.* 120, 1635–1651. doi: 10.1007/s00122-010-1282-0
- Santalla, M., Monteagudo, A. B., Gonzalez, A. M., and De Ron, A. M. (2004). Agronomical and quality traits of runner bean germplasm and implications for breeding. *Euphytica* 135, 205–215. doi: 10.1023/B:EUPH.0000014912.07993.e7
- Santalla, M., Rodiño, A. P., and De Ron, A. M. (2002). Allozyme evidence supporting southwester Europe as a secondary center of genetic diversity for common bean. *Theor. Appl. Genet.* 104, 934–944. doi: 10.1007/s00122-001-0844-6
- Schinkel, C., and Gepts, P. (1988). Phaseolin diversity in the tepary bean *Phaseolus acutifolius* A. Gray. *Plant Breed.* 101, 292–301. doi: 10.1111/j.1439-0523.1988.tb00301.x
- Schinkel, C., and Gepts, P. (1989). Allozyme variability in the tepary bean, *Phaseolus acutifolius* A. Gray. *Plant Breed.* 102, 182–195. doi: 10.1111/j.1439-0523.1989.tb00336.x
- Schmit, V., and Debouck, D. G. (1991). Observations on the origin of *Phaseolus polyanthus* Greenman. *Econ. Bot.* 45, 345–364. doi: 10.1007/BF02887077
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Serrano-Serrano, M. L., Andueza-Noh, R. H., Martínez-Castillo, J., Debouck, D. G., and Chacón, M. I. (2012). Evolution and domestication of Lima bean (*Phaseolus lunatus* L.) in Mexico: Evidence from ribosomal DNA. *Crop Sci.* 52, 1698–1712. doi: 10.2135/cropsci2011.12.0642
- Serrano-Serrano, M. L., Hernandez-Torres, J., Castillo-Villamizar, G., Debouck, D. G., and Chacón, M. I. (2010). Gene pools in wild Lima bean (*Phaseolus lunatus* L.) from the Americas: evidences for an Andean origin and past migrations. *Mol. Phylogenet. Evol.* 54, 76–87. doi: 10.1016/j.ympev.2009.08.028
- Shii, C. T., Rabakoarihanta, A., Mok, M. C., and Mok, D. W. S. (1982). Embryo development in reciprocal crosses of *Phaseolus vulgaris* and *Phaseolus coccineus*. *Theor. Appl. Genet.* 62, 59–64. doi: 10.1007/BF00276284
- Sicard, D., Nanni, L., Porfiri, O., Bulfon, D., and Papa, R. (2005). Genetic diversity of *Phaseolus vulgaris* L. and *P. coccineus* L. landraces in central Italy. *Plant Breed.* 124, 464–472. doi: 10.1111/j.1439-0523.2005.01137.x
- Siebe, C. (2000). Age and archaeological implications of xitle volcano, southwestern basin of Mexico-city. *J. Volcanol. Geotherm. Res.* 104, 45–64. doi: 10.1016/S0377-0273(00)00199-2
- Siebe, C., Rodríguez-Lara, V., Schaaf, P., and Abrams, M. (2004). Radiocarbon ages of Holocene Pelado, Guespalapa, and Chichinautzin scoria cones, south of Mexico City: implications for archaeology and future hazards. *Bull. Volcanol.* 66, 203–225. doi: 10.1007/s00445-003-0304-z
- Sofi, P. A., Rana, J. C., and Bhat, N. A. (2014). Pattern of variation in common bean (*Phaseolus vulgaris* L.) genetic resources of Jammu and Kashmir. *J. Food Legumes* 27, 197–201.
- Spataro, G., Tiranti, B., Arcaleni, P., Bellucci, E., Attene, G., Papa, R., et al. (2011). Genetic diversity and structure of a worldwide collection of *Phaseolus coccineus* L. *Theor. Appl. Genet.* 122, 1281–1291. doi: 10.1007/s00122-011-1530-y
- Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M. B., Ross-Ibarra, J., Myers, C. L., et al. (2012). Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11878–11883. doi: 10.1073/pnas.1201961109
- Tanksley, S. D., and McCouch, S. R. (1997). Seed banks and molecular maps: unlocking genetic potential from the wild. *Science* 277, 1063–1066. doi: 10.1126/science.277.5329.1063
- Tarrago, M. N. (1980). “El proceso de agriculturización en el Noroeste Argentino, zona Valliserrana,” in *Actas del V Congreso Nacional Arqueología Argentina*, Vol. 1, (San Juan: Universidad de San Juan), 181–217.
- Toro, O., Tohme, J., and Debouck, D. G. (1990). *Wild Bean (Phaseolus vulgaris L.): Description and Distribution*. Cali: Centro Internacional de Agricultura Tropical.
- van Heerwaarden, J., Doebley, J., Briggs, W. H., Glaubitz, J. C., Goodman, M. M., Sanchez-Gonzalez, J. J., et al. (2011). Genetic signals of origin, spread, and introgression in a large sample of maize landraces. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1088–1092. doi: 10.1073/pnas.1013011108
- Vaz Patto, M. C., Amarowicz, R., Aryee, A. N., Boye, J. I., Chung, H. J., Martín-Cabrejas, M. A., et al. (2015). Achievements and challenges in improving the nutritional quality of food legumes. *Crit. Rev. Plant Sci.* 34, 105–143. doi: 10.1080/07352689.2014.897907

- Vitte, C., Ishii, T., Lamy, F., Brar, D., and Panaud, O. (2004). Genomic paleontology provides evidence for two distinct origins of Asian rice (*Oryza sativa* L.). *Mol. Genet. Genom.* 272, 504–511. doi: 10.1007/s00438-004-1069-6
- Vlasova, A., Capella-Gutiérrez, S., Rendón-Anaya, M., Hernández-Oñate, M., Minoche, A. E., Erb, I., et al. (2016). Genome and transcriptome analysis of the Mesoamerican common bean and the role of gene duplications in establishing tissue and temporal specialization of genes. *Genome Biol.* 17, 32. doi: 10.1186/s13059-016-0883-6
- vonHoldt, B. M., Pollinger, J. P., Lohmueller, K. E., Han, E., Parker, H. G., Quignon, P., et al. (2010). Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464, 898–902. doi: 10.1038/nature08837
- Vrebalov, J., Panb, I. L., Arroyo, A. J., McQuinn, R., Chung, M., Poole, M., et al. (2009). Fleshy fruit expansion and ripening are regulated by the tomato *SHATTERPROOF* gene *TAGL1*. *Plant Cell* 21, 3041–3062. doi: 10.1105/tpc.109.066936
- Wall, J. R. (1970). Experimental introgression in the genus *Phaseolus*. 1. Effect of mating systems on interspecific gene flow. *Evolution* 24, 356–366. doi: 10.2307/2406810
- Wang, I. J., Glor, R. E., and Losos, J. B. (2013). Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* 16, 175–182. doi: 10.1111/ele.12025
- Westphal, E. (ed.) (1974). “*Phaseolus*,” in *Pulses in Ethiopia, their Taxonomy and Agricultural Significance*, (Wageningen: Centre for Agricultural Publishing and Documentation, (PUDOC)), 129–176.
- Wortmann, C. S., Kirkby, R. A., Eledu, C. A., and Allen, D. J. (1998). *Atlas of Common Bean (Phaseolus vulgaris L.) Production in Africa*. Cali: Centro Internacional de Agricultura Tropical (CIAT), 133.
- Xia, Q., Guo, Y., Zhang, Z., Li, D., Xuan, Z., Li, Z., et al. (2009). Complete resequencing of 40 genomes reveals domestication events and genes in silkworm (*Bombyx*). *Science* 326, 433–436. doi: 10.1126/science.1176620
- Xu, X., Liu, X., Ge, S., Jensen, J. D., Hu, F., Li, X., et al. (2012). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111. doi: 10.1038/nbt.2050
- Young, K. R., Ulloa Ulloa, C., and Luteyn, J. L. (2002). Plant evolution and endemism in Andean South America: an introduction. *Bot. Rev.* 68, 4–21. doi: 10.1663/0006-8101(2002)068[0004:PEAEIA]2.0.CO;2
- Zhang, X., Blair, M. W., and Wang, S. (2008). Genetic diversity of Chinese common bean (*Phaseolus vulgaris* L.) landraces assessed with simple sequence repeats markers. *Theor. Appl. Genet.* 117, 629–640. doi: 10.1007/s00122-008-0807-2
- Zizumbo-Villarreal, D., and Colunga-GarcíaMarín, P. (2010). Origin of agriculture and plant domestication in West Mesoamerica. *Genet. Res. Crop Evol.* 57, 813–825. doi: 10.1007/s10722-009-9521-4
- Zizumbo-Villarreal, D., Papa, R., Hufford, M., Repinski, S., and Gepts, P. (2009). Identification of a new wild population of *Phaseolus vulgaris* L. in western Jalisco, Mexico, near the Mesoamerican domestication center of common bean. *BIC* 52, 24–25.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bitocchi, Rau, Bellucci, Rodriguez, Murgia, Gioia, Santo, Nanni, Attene and Papa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genetic Diversity of Croatian Common Bean Landraces

Klaudija Carović-Stanko^{1,2*}, Zlatko Liber^{2,3}, Monika Vidak¹, Ana Barešić¹,
Martina Grdiša^{1,2}, Boris Lazarević⁴ and Zlatko Šatović^{1,2}

¹ Department of Seed Science and Technology, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia, ² Centre of Excellence for Biodiversity and Molecular Plant Breeding (CroP-BioDiv), Zagreb, Croatia, ³ Department of Botany, Faculty of Science, University of Zagreb, Zagreb, Croatia, ⁴ Department of Plant Nutrition, Faculty of Agriculture, University of Zagreb, Zagreb, Croatia

OPEN ACCESS

Edited by:

Nicolas Rispaï,
Consejo Superior de Investigaciones
Científicas, Spain

Reviewed by:

Peter C. Mckeown,
NUI Galway, Ireland
Roberto Papa,
Università Politecnica delle Marche,
Italy
Elena Bitocchi,
Università Politecnica delle Marche,
Italy

*Correspondence:

Klaudija Carović-Stanko
kcarovic@agr.hr

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 30 December 2016

Accepted: 03 April 2017

Published: 20 April 2017

Citation:

Carović-Stanko K, Liber Z, Vidak M,
Barešić A, Grdiša M, Lazarević B and
Šatović Z (2017) Genetic Diversity of
Croatian Common Bean Landraces.
Front. Plant Sci. 8:604.
doi: 10.3389/fpls.2017.00604

In Croatia, the majority of the common bean production is based on local landraces, grown by small-scale farmers in low input production systems. Landraces are adapted to the specific growing conditions and agro-environments and show a great morphological diversity. These local landraces are in danger of genetic erosion caused by complex socio-economic changes in rural communities. The low profitability of farms and their small size, the advanced age of farmers and the replacement of traditional landraces with modern bean cultivars and/or other more profitable crops have been identified as the major factors affecting genetic erosion. Three hundred accessions belonging to most widely used landraces were evaluated by phaseolin genotyping and microsatellite marker analysis. A total of 183 different multi-locus genotypes in the panel of 300 accessions were revealed using 26 microsatellite markers. Out of 183 accessions, 27.32% were of Mesoamerican origin, 68.31% of Andean, while 4.37% of accessions represented putative hybrids between gene pools. Accessions of Andean origin were further classified into phaseolin type II (“H” or “C”) and III (“T”), the latter being more frequent. A model-based cluster analysis based on microsatellite markers revealed the presence of three clusters in congruence with the results of phaseolin type analysis.

Keywords: common bean, landrace, origin, phaseolin type, microsatellite markers

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is a valuable legume for human consumption worldwide, being an important source of high quality proteins, carbohydrates, vitamins, minerals, dietary fiber, phytonutrients (flavonoids, lignins, phytosterols) and antioxidants (Cardador-Martínez et al., 2002; Reynoso-Camacho et al., 2006). Many of these compounds have important beneficial effects on human health, therefore, common bean has considerable potential as a functional food.

Common bean was introduced into Europe from mutually independent domestication centers, Central and South America, where the Mesoamerican and the Andean cultivated gene pools originated (Gepts and Debouck, 1991). Common bean landraces originated from these two gene pools were introduced into Europe at different times. The Mesoamerican common bean landraces probably arrived in Europe through Spain and Portugal in 1,506, and the Andean in the same way in 1,528, after the exploration of Peru by Pizarro (Gioia et al., 2013). Subsequent spread of common bean landraces throughout Europe was very complex with several introductions from various regions of the Americas, combined with frequent exchanges between European and other

Mediterranean countries (Papa et al., 2006). The common bean is distributed in Europe, Asia and Africa, where it presents similarities to Andean and Mesoamerican gene pools or forms hybrids between both gene pools (Chávez-Servia et al., 2016). Gene flow between domesticated and wild beans led to substantial introgression of alleles from the domesticated gene pool into the wild gene pool and vice versa (Pathania et al., 2014). Gene pool diversity has been validated using various marker systems including seed size, plant morphology, phaseolin seed protein patterns, allozymes and molecular markers (Asfaw et al., 2009). Europe, Brazil, central-eastern and southern Africa and China have been suggested as the secondary centers of diversification for common bean (Bellucci et al., 2014).

It has been proven that phaseolin, the major seed storage protein of common bean, is an important molecular marker in studies of genetic diversity and evolution of common bean populations due to its functional and structural properties (De la Fuente et al., 2012). Two gene pools of domestication are distinguished in the species and are characterized by morphological differences as well as by phaseolin types. Predominant phaseolyn types are “S” (Mesoamerican) or “T”/“C”/“H” (Andean), (Raggi et al., 2013). According to studies based on phaseolin analyses the Andean gene pool of common bean is always prevalent in Europe, between 66 and 76% (Lioi, 1989; Logozzo et al., 2007; Angioi et al., 2010).

In Europe in recent decades, in response to market demands, landraces have progressively been replaced by improved cultivars but some studies have shown that many landraces survive on-farm in marginal areas of several European countries (Lioi et al., 2005). In Croatia the production of common bean is based on landraces which are adapted to the specific growing conditions and agro-environments which display high levels of morphological diversity (Čupić et al., 2012). Landraces are traditionally grown in low-input production systems. Preservation of the genetic diversity that is held by small-scale farmers could provide important sources of genetic resistance for plant breeders, as they are likely to contain alleles for local adaptations, disease resistance, and tolerance to the principal climate adversities in the region. However, in recent years, landraces are in danger of genetic erosion caused by complex socio-economic changes in rural communities (the low profitability of farms, their small size, and the advanced age of farmers, the replacement of traditional landraces with modern bean cultivars and/or other more profitable crops; FAO, 2008).

In order to depict the origin and diversity of common bean landraces, it is necessary to conduct analyses at the morphological and genetic level. Therefore, the aim of this research was the assessment of genetic diversity and structure of Croatian common bean landraces using microsatellites and the determination of their origin by phaseolin marker analysis. The results were discussed in a broader, European context.

MATERIALS AND METHODS

Plant Material

Three-hundred accessions of common bean landraces were collected from diverse geographical regions of Croatia (Figure 1)

and grown in unreplicated field plots at the experiment field in Maksimir, Zagreb (45.8293 N, 16.0334 E) in year 2014. They were classified as determinate (type I growth habit; Singh, 1982) or indeterminate. After preliminary analyses, accessions showing similar seed color/pattern and habit while collected from the same location (village) but from different households were excluded. Finally, 183 accessions were chosen and a single plant of each accession was used in the analyses. A list of the accessions, along with their “passport” information, as well as the information on habit, phaseolin type and cluster (based on model-based clustering methods, see below) is available in Table S1. Some examples of seed color/pattern diversity of accessions are shown in Figure S1. The accessions are held at the Department of Seed Science and Technology, Faculty of Agriculture, University of Zagreb.

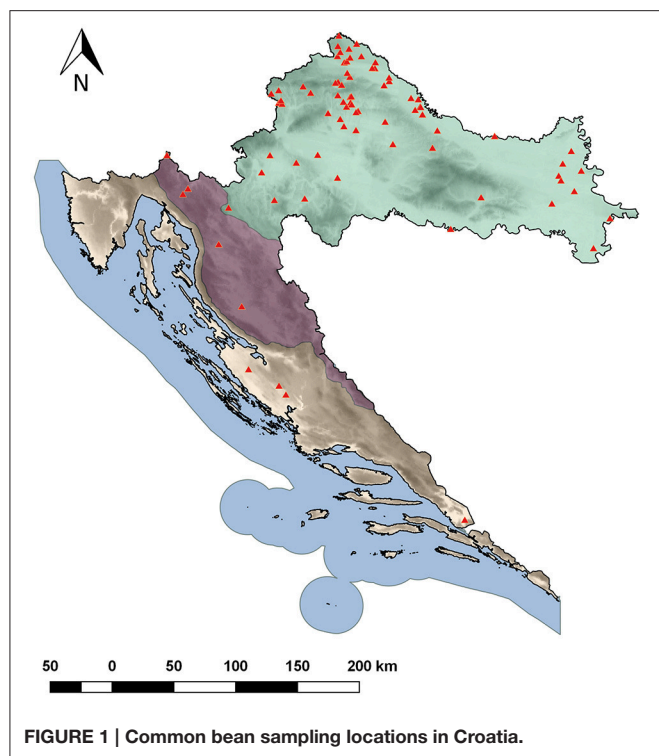
DNA Extraction, Phaseolin, and Microsatellite Analysis

Using Plant DNeasy 96 kit (Qiagen®), DNA was isolated from 25 mg of silica-gel dried leaves according to the manufacturer's instructions without any additional clean-up. Tailed PCR approach (Schuelke, 2000) was used for amplification of phaseolin sequences (Kami et al., 1995). The 20 µl of the PCR mix contained 2 pmol of the tailed forward primer (5'-TGTAACACGACGGCCAGTAGCATATTCTAGAGGCTCC-3'), 8 pmol of reverse (5'-GCTCAGTTCCTCAATCTGTTC-3'), 8 pmol of FAM labeled M13 primer (5'-TGTAACACGACGGCCAGT-3'), 1 × PCR buffer, 4 pmol of each dNTP, 0.5 U TaqTM HS DNA Polymerase (Takara Bio Inc.) and 5 ng of template DNA. PCR protocol with an initial touchdown cycles (94°C for 5 min; 5 cycles of 45 s at 94°C, 30 s at 60°C, which was lowered by 1°C in each cycle, and 90 s at 72°C; 25 cycles of 45 s at 94°C, 30 s at 55°C, and 90 s at 72°C; and 8-min extension step at 72°C) was employed (Radosavljević et al., 2011). The PCR products were sent to GeneScan service MacroGen® (South Korea) where they were detected on an ABI 3730xL DNA analyzer (Applied Biosystems®) by and analyzed GeneMapper 4.0 computer program (Applied Biosystems®).

Twenty-six PCR primer pairs were used for microsatellite analysis (Table S2). DNA amplification was performed using multiplex PCR mix and the same two-step PCR protocol with an initial touchdown cycle as in phaseolin type determination. The 20 µl of PCR mix contained 5 pmol of each of four fluorescent labeled forward primers (6-FAM, VIC, NED, PET; Applied Biosystems®), 5 pmol of reverse primers, 1 × PCR buffer, 4 pmol of each dNTP, 0.5 U TaqTM HS DNA Polymerase (Takara Bio Inc.) and 5 ng of template DNA. Fluorescent labeled PCR products were detected on an ABI 3730XL (Applied Biosystems®) by GeneScan service (MacroGen®). Allele sizes (in base pairs) of PCR products were estimated using GeneMapper 4.0 computer program (Applied Biosystems®).

Data Analysis

The average number of alleles per locus (N_a), observed heterozygosity (H_o), and gene diversity (expected heterozygosity; H_E) for each microsatellite locus was calculated in GENEPOP 4.0 (Raymond and Rousset, 1995).



The proportion-of-shared-alleles distance (Bowcock et al., 1994) between pairs of accessions genotyped using 26 microsatellites was calculated using MICROSAT (Minch et al., 1997). Cluster analysis was performed using the Fitch-Margoliash least-squares algorithm in PHYLIP (Felsenstein, 2004). The reliability of the tree topology was assessed via bootstrapping (Felsenstein, 1985) over 1,000 replicates generated by MICROSAT and subsequently used in PHYLIP.

A model-based clustering method was applied on multilocus microsatellite data to infer genetic structure and define the number of clusters in the dataset using the software STRUCTURE ver. 2.3.3 (Pritchard et al., 2000). Thirty runs per each cluster (K) ranging from 1 to 11 were carried out on the Isabella computer cluster at the University of Zagreb, University Computing Centre (SRCE). Each run consisted of a burn-in period of 200,000 steps followed by 10^6 MCMC (Monte Carlo Markov Chain) replicates assuming admixture model and correlated allele frequencies. No prior information was used to define the clusters. The choice of the most likely number of clusters (K) was carried out by comparing the average estimates of the likelihood of the data, $\ln[Pr(X|K)]$, for each value of K (Pritchard et al., 2000), as well as by calculating an ad hoc statistic ΔK , based on the rate of change in the log probability of data between successive K -values as described by Evanno et al. (2005). The program STRUCTURE HARVESTER v0.6.92 was used to process the STRUCTURE results files (Earl and von Holdt, 2012). Runs were clustered and averaged using CLUMPAK (Kopelman et al., 2015). The accessions were assigned to a particular cluster if an arbitrary value of $Q > 75\%$ of their genome was estimated

to belong to that cluster (Matsuoka et al., 2002), while those accessions having the membership probabilities $Q < 75\%$ for each cluster were considered as of “mixed origin.” The likelihood-ratio chi-square test in SAS (SAS Institute, 2004) was used to test for dependence between phaseolin type and cluster membership of the accessions. The strength of association was assessed by calculating Cramér’s V , the measure that reaches the maximum value of 1 when the two variables (i.e., classification criteria) are equal to each other.

Genetic diversity of Croatian common bean accessions was analyzed by classifying the accessions into two and into three germplasm groups according to the results of model-based cluster analyses of 26 microsatellite loci and phaseolin type. The accessions considered as of “mixed origin” as well as those that did not show correspondence between phaseolin type and cluster membership in STRUCTURE were excluded from further analysis.

The genetic diversity of each group of accessions was assessed by calculating the average number of alleles per locus (N_a) and allelic richness (N_{ar}) in FSTAT (Goudet, 2002) as well as gene diversity (H_E) in GENEPOP 4.0 (Raymond and Rousset, 1995). In order to compare the values of allelic richness (N_{ar}) and gene diversity (H_E) between/among groups, the repeated measures analysis of variance was carried out using PROC GLM in SAS followed by *post-hoc* Bonferroni’s adjustments.

The analysis of molecular variance (AMOVA; Excoffier et al., 1992) using ARLEQUIN ver. 3.0 (Excoffier et al., 2005) was used to partition the total microsatellite diversity between and within groups of accessions. The variance components were tested statistically by non-parametric randomization tests using 10,000 permutations.

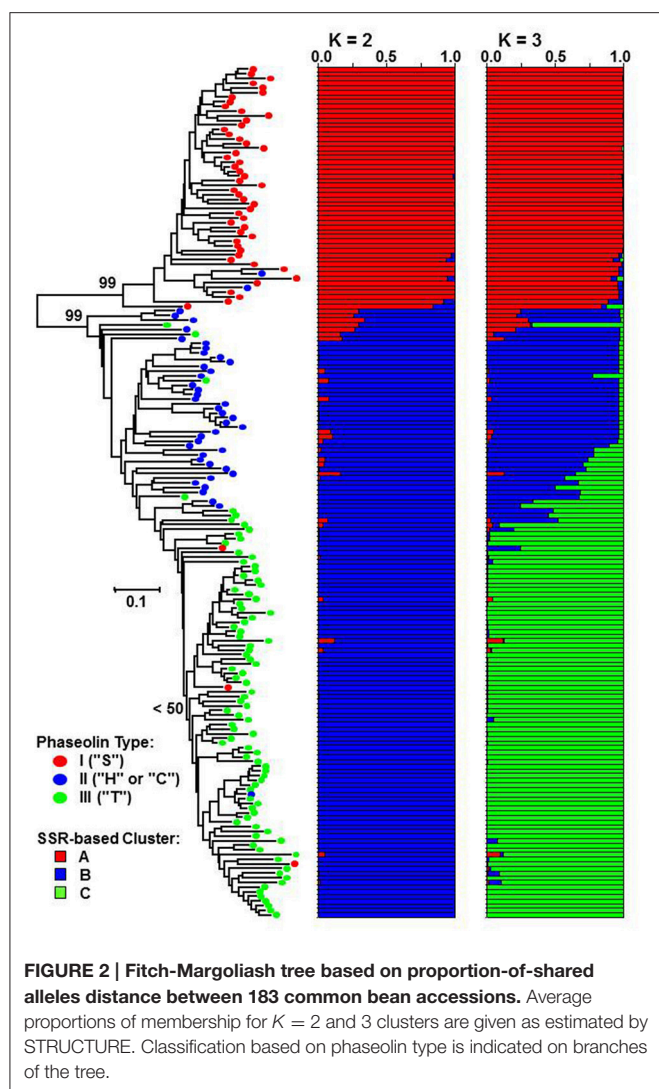
RESULTS

Classification of Croatian Common Bean Accessions

Phaseolin analysis of 183 accessions revealed that 53 (28.96%) of them were of phaseolin type I (Mesoamerican; “S”), 42 (22.95%) of II (Andean; “H” or “C”) and 88 (48.09%) of III (Andean; “T”).

A total of 137 alleles were detected at 26 SSR loci, ranging from 2 (BMb469, BMd12, BMd22, BMd25, BMd45, BMd46, and BMd47) to 19 (PVat007) alleles per microsatellite locus with an average of 5.27 (Table S2). Observed heterozygosity (H_O) values of all the markers were equal to zero, i.e., all the samples were completely homozygous for all the loci. Average gene diversity (H_E) was $H_E = 0.572$, ranging from 0.389 (BMd12) to 0.885 (PVat007).

The genetic distance between pairs of accessions based on the proportion-of-shared-alleles distance measure ranged from $D_{psa} = 0.038$ (50 common alleles out of 52) to $D_{psa} = 1.000$ (no alleles in common) with the average of $D_{psa} = 0.577$. The Fitch-Margoliash tree grouped the accessions into two well-supported clades (bootstrap support value: 99%) corresponding to Mesoamerican and Andean origin of accessions as identified by phaseolin analysis (Figure 2). The subclade containing the great majority of phaseolin type III accessions could be identified within the Andean clade, although the monophyly



of that group was not supported by a bootstrap value higher than 50%.

The STRUCTURE analysis, as expected, identified $K = 2$ as the most likely number of clusters ($\Delta K = 21410.10$). The most of the accessions of Mesoamerican origin (phaseolin type I) were assigned to cluster A, while the accession of Andean origin (phaseolin type II and III) were assigned to cluster B. The association between phaseolin type and cluster membership was highly significant and nearly complete ($\chi^2 = 174.72$; $df = 1$; $P < 0.0001$; Cramér's $V = 0.93$). At $K = 3$, the newly formed cluster (cluster C) clearly grouped the phaseolin type III accessions that corresponded to the subclade in the Fitch-Margoliash tree ($\chi^2 = 281.97$; $df = 4$; $P < 0.0001$; Cramér's $V = 0.89$).

At $K = 2$, a total of five accessions could be considered as of “mixed origin” having the membership probabilities $Q < 75\%$ for both clusters. Furthermore, additional five accessions did not show the correspondence between phaseolin type and the membership according to model-based clustering analysis based on microsatellite loci (i.e., Mesoamerican group: cluster A/phaseolin type I; Andean group: cluster B phaseolin type II or

III). Thus, a total of 10 accessions (5.46%) could be considered as putative hybrids between gene pools and were excluded from the subsequent analyses of genetic diversity.

At $K = 3$, a total of 16 accessions were classified as of “mixed origin,” while 13 accessions did not show the correspondence (i.e., Mesoamerican group: cluster A/phaseolin type I, Andean group B: cluster B/phaseolin type II; Andean group C: cluster C/phaseolin type III). As four accessions were classified as both of “mixed origin” and “non-corresponding,” a total of 25 accessions (13.66%) originating from hybridization among the three groups were excluded from the subsequent analyses of genetic diversity. Finally, the classification of 183 Croatian common bean accessions would be the following: (1) 50 accessions (27.32%) belonged to Mesoamerican group, 27 (14.75%) to Andean group B and 81 (44.26%) to Andean group C; (2) Four (2.19%) accessions were putative hybrids between Mesoamerican group and Andean group B, another four (2.19%) were hybrids between Mesoamerican and Andean group C, while 17 (9.29%) accessions were hybrids between Andean groups B and C.

Genetic Diversity of Germplasm Groups

By classifying accessions into two groups [Mesoamerican group: cluster A at $K = 2$ /phaseolin type I vs. Andean group: cluster B at $K = 2$ /phaseolin type II or III], the Andean group of accessions showed slightly higher values of allelic richness (N_{ar}) as well as gene diversity (H_E) than the Mesoamerican group but the differences were not significant following the analysis of variance (Table 1). Average genetic distance between pairs of accessions belonging to Mesoamerican group ($D_{psa} = 0.277$) was lower than in Andean group ($D_{psa} = 0.356$) while the average distance between pairs belonging to different groups was considerably higher ($D_{psa} = 0.887$). The analysis of molecular variance (AMOVA) revealed that that 63.34% of microsatellite diversity could be attributed to differences between groups ($\Phi_{ST} = 0.633$; $P < 0.0001$; Table 2). Mesoamerican group of accessions consisted mostly of accessions of indeterminate growth habit (43 out of 50) while in the Andean group the accessions of determinate growth habit (85 out of 123). The association between group membership and growth habit was highly significant, but moderate ($\chi^2 = 46.54$; $df = 1$; $P < 0.0001$; Cramér's $V = 0.50$).

The diversity analysis of three groups (Mesoamerican group: cluster A at $K = 3$ /phaseolin type I, Andean group B: cluster B at $K = 3$ /phaseolin type II; Andean group C: cluster C at $K = 3$ /phaseolin type III) revealed that the Mesoamerican group had the highest allelic richness (N_{ar}) while the Andean group B had the highest gene diversity (H_E). The lowest values of both measures were found in the Andean group C. However, the differences among groups were not significant (Table 1). The AMOVA analysis based on three groups revealed the similar results as in case of classification into two groups: 64.44% of diversity was attributed to differences between groups ($\Phi_{ST} = 0.644$; $P < 0.0001$; Table 2). The lowest pairwise Φ_{ST} value between groups types was found between two Andean groups ($\Phi_{ST(B/C)} = 0.420$) while the Φ_{ST} -values between Mesoamerican group and Andean group B as well as between Mesoamerican group and Andean group C were considerably higher ($\Phi_{ST(A/B)}$

TABLE 1 | Genetic diversity of Croatian common bean accessions based on 26 microsatellite loci.

Cluster/Phaseolin type	N _{Acc}	S	T	N _a	N _{ar}	N _{pr}	H _E
A/I (Mesoamerican)	50	7	43	3.269	3.269	31	0.277
B/II or III (Andean)	123	85	38	4.038	3.677	51	0.356
P-value					0.455		0.249
A/I (Mesoamerican)	50	7	43	3.269	3.100	31	0.277
B/II (Andean)	27	0	27	2.885	2.885	12	0.314
C/III (Andean)	81	77	4	2.846	2.538	18	0.263
P-value					0.291		0.679

Accessions were classified into two and three groups according to model-based cluster analysis and phaseolin type. N_{Acc}, No. of Accessions; S, No. of Accessions with Determinate Growth Habit; T, No. of Accessions with Indeterminate Growth Habit; N_a, No. of Alleles; N_{ar}, Allelic Richness; N_{pr}, No. of Private Alleles; H_E, Gene Diversity; P-value, significance level of the F-test.

TABLE 2 | Analysis of molecular variance for the partitioning of microsatellite diversity of Croatian common bean accessions classified into (A) two as well as (B) three groups according to model-based cluster analysis and phaseolin type.

Analysis	Source of variation	df	Variance components	% Total variance	Φ-statistic	P(Φ)
(A)	Between groups	1	7.443	63.34	0.633	<0.0001
	Within groups	344	4.308	36.66		
(B)	Among groups	2	6.433	64.44	0.644	<0.0001
	Within groups	313	3.550	35.56		

P(Φ)–Φ -statistic probability level after 10,000 permutations.

= 0.654; $\Phi_{ST/(A/C)} = 0.706$). All the pairwise Φ_{ST} -values were highly significant ($P < 0.0001$). The same pattern was found by analyzing the average genetic distance between pairs of accessions belonging to different groups. The average distance between Andean groups A and B ($D_{psa(B/C)} = 0.481$) was substantially lower than the distances between Mesoamerican group and Andean groups ($D_{psa(A/B)} = 0.831$; $D_{psa(A/C)} = 0.908$). As already mentioned, the Mesoamerican group of accessions consisted mostly of accessions of indeterminate growth habit (43 out of 50). Moreover, the Andean group B included exclusively the accession of indeterminate growth habit, while the great majority of accessions belonging to Andean group C was of determinate growth habit (77 out of 81) leading to a strong association between group membership and growth habit ($\chi^2 = 146.04$; $df = 2$; $P < 0.0001$; Cramér's $V = 0.87$). Thus, the estimated probability of correct prediction of growth habit based on phaseolin type of accession was $P = 0.93$.

DISCUSSION

Origin of Croatian Common Bean Germplasm

To determine the evolutionary origin of Croatian common bean accessions we combined the results of phaseolin marker analysis and microsatellite genotyping. Out of 183 accessions, 27.32% are of Mesoamerican origin, 68.31% of Andean, while 4.37% of accessions represent putative hybrids between gene pools. Our results are in line with the findings of numerous previous studies that the European common bean germplasm originates from both gene pools, Mesoamerican and Andean, later being more

frequently found (see Bellucci et al., 2014 for a review). The proportion of landraces of the Mesoamerican origin tends to increase in eastern and south-eastern Europe as shown, e.g., in case of Albania (Logozzo et al., 2007), Bulgaria (Svetleva et al., 2006), Macedonia (Maras et al., 2015) and Greece (Lioi, 1989). However, Maras et al. (2015) reported that the proportions found in accessions from Bosnia and Herzegovina, Croatia, Serbia and Slovenia were very similar to those found in the Iberian Peninsula and Italy indicating that common bean was introduced into the western Balkans mainly from the Mediterranean Basin. In case of Croatia these findings have been confirmed by the present study that included substantially more accessions. However, in contrast to Italian and Spanish common bean germplasm in which the Andean phaseolin type is “C” prevails over the type “T” (Logozzo et al., 2007; Angioi et al., 2010; Raggi et al., 2013), “T” (i.e., Andean group C: cluster C at $K = 3$ /phaseolin type III) is clearly the most common phaseolin type found in Croatian germplasm as in most other European countries (Logozzo et al., 2007).

Genetic Diversity of the Mesoamerican and Andean Group of Accessions

Nearly complete correspondence of classifications based on phaseolin analysis and model-based clustering using microsatellite markers as well as a strong association between group membership and growth habit in Croatian common bean germplasm could be explained by a series of sequential bottlenecks during domestication, early introduction to Portugal and Spain and eastward expansion throughout Europe. The Mesoamerican group of accessions as well as the Andean group B consists mostly of accessions of indeterminate growth habit

while the great majority of accessions belonging to Andean group C have determinate growth habit. A strong association between phaseolin pattern and growth habit was reported by Raggi et al. (2013) by analyzing Italian common bean landraces: plants with climbing ability were prevalent in the “C” (Mesoamerican group) and “S” (Andean group B) phaseolin pattern groups, while bush plants were prevalent in the “T” (Andean group C) group. Moreover, Kwak et al. (2012) reported that determinate types were found mainly in Andean subpopulations.

Similar levels of gene diversity of Croatian common bean accessions of Andean origin as compared with those of Mesoamerican origin are in line with findings previously reported by Santalla et al. (2002) in Iberian landraces as well as by Angioi et al. (2010) at European level, while in the domestication centers, the diversity observed in Mesoamerican gene pool is higher than in the Andean (Kwak and Gepts, 2009). Angioi et al. (2010) offered two plausible, and not mutually exclusive, explanations: (a) further selection in Europe might have reduced the variation of the Mesoamerican germplasm, and/or (b) diversity of Mesoamerican introductions to Europe was already reduced when compared with the Mesoamerican gene pool. Additionally, the apparent incongruence can be explained by the fact that the Andean gene pool is represented by two separate groups of accessions (determinate/indeterminate) that resulted from divergent selection during domestication in the Andes. It is well-documented that the divergence between Mesoamerican and Andean gene pools preceded the domestication that occurred independently in two geographic regions (Kwak and Gepts, 2009). The wild Andean gene pool diverged from the wild Mesoamerican gene pool, the origin of the species (Bitocchi et al., 2012), with a strong bottleneck, as shown by considerably lower genetic diversity of wild Andean gene pool in comparison to Mesoamerican (Schmutz et al., 2014). Wild common beans are all indeterminate and it would be reasonable to assume that the first domesticated types were also indeterminate as determinacy is a trait selected during or after domestication (Kwak et al., 2012). In Mesoamerican domestication center, indeterminate types were the valuable component of traditional maize-bean-squash multicrop system (Zizumbo-Villarreal and Colunga-García, 2010). On the other hand, in Andean domestication center, early farmers did not have a suitable crop that could serve as a physical support for viny common bean as domestication of common bean preceded the introduction of maize (Kwak et al., 2012). By constant selection of genotypes displaying a more compact growth habit, determinacy was included in the group of traits known as the domestication syndrome (Hammer, 1984; Koinange et al., 1996). Indeterminate landraces were not abandoned, and after maize introduction, maize/bean intercropping has gained importance and remained until nowadays (Lithourgidis et al., 2011). Although the majority of determinate type accessions originate from the Andean gene pool, the determinacy has been selected independently in both domestication centers (Kwak et al., 2012). The process of domestication led to reduction of genetic diversity, but, interestingly, the bottleneck effect was threefold greater in Mesoamerica as compared to the Andes as shown by Bitocchi et al. (2013). Bearing in mind that the

wild Andean gene pool was strongly impoverished as a result of a bottleneck that occurred before domestication, Bitocchi et al. (2013) concluded that it would be the reason why the subsequent domestication bottleneck had the minor effect in the Andes. Another possible explanation would be that the divergent selection (indeterminate/determinate types) during domestication that was prominent in the Andes but presumably negligible in the Mesoamerican region. While in Mesoamerica the reduction in diversity of neutral genes followed the reduction of the genes under selection, divergent selection in the Andes maintained or even increased diversity by possible effects of introgression/interchange during improvement (Burger et al., 2008).

Hybridization between Mesoamerican and Andean Gene Pools

Out of 183 accessions, a total of 25 (13.66%) were classified as putative hybrids: Four (2.19%) accessions were putative hybrids between Mesoamerican group and Andean group B, another four (2.19%) were hybrids between Mesoamerican and Andean group C, while 17 (9.29%) accessions were hybrids between Andean groups B and C. Thus, the proportion of hybrids between Mesoamerican and Andean gene pools amounts to only 4.37%. This result is in striking disagreement with the results reported by, *inter alia*, Angioi et al. (2010) and Gioia et al. (2013). By analyzing 307 European common bean accessions they estimated that about 44% of them were derived from hybridization between Mesoamerican and Andean gene pools. Similarly, Gioia et al. (2013) reported that 40.2% of 256 European common bean accessions derived from hybridization between two gene pools. Moreover, Angioi et al. (2010) showed that the hybridization was particularly frequent in Central Europe as compared to Italy and the Iberian Peninsula. Our results show a very different picture with much less frequent hybridization compared to Angioi et al. (2010) and Gioia et al. (2013). Thus, in Croatian common bean landraces inter-gene pool hybridization appears to have been very limited, even if differences in the methods used to detect the hybrids should also be considered when comparing our results to that obtained in the rest of Europe. Indeed, Angioi et al. (2010) carried out the analysis by combining chloroplast microsatellites and two nuclear loci (for phaseolin types and *Pv-shatterproof1*), and Gioia et al. (2013) used nuclear and chloroplast microsatellites as well as two nuclear loci (for phaseolin types and *Pv-shatterproof1*), while this study was based on nuclear microsatellites and phaseolin marker.

CONCLUSION

This study provides a comprehensive picture of genetic diversity and structure of Croatian common bean germplasm. Out of 183 accessions, 27.32% were of Mesoamerican origin, 68.31% of Andean, while 4.37% of accessions represented putative hybrids between gene pools. For the most part, the classification of common bean accessions according to phaseolin type analysis was in congruence with the results of both distance-based and model-based analyses of microsatellite marker data. The

Mesoamerican group (cluster A/phaseolin type I) of accessions as well as the Andean group B (cluster B/phaseolin type II) consisted mostly of accessions of indeterminate growth habit while the great majority of accessions belonging to Andean group C (cluster C/phaseolin type III) had determinate growth habit. Nearly complete correspondence of classifications based on phaseolin analysis and microsatellite markers as well as a strong association between group membership and growth habit in Croatian common bean germplasm could be explained by a series of sequential bottlenecks.

AUTHOR CONTRIBUTIONS

Conceived and designed the manuscript: KC, MV, and ZŠ. Contributed to analysis: ZL, MV, AB, and MG. Analyzed the data:

ZL and ZŠ. Wrote the manuscript: KC, ZL, MV, AB, MG, BL, and ZŠ.

ACKNOWLEDGMENTS

This work has been fully supported by Croatian Science Foundation under the project UIP-11-2013-3290 Genetic basis of bioactive nutrient content in Croatian common bean landraces (BeanQual).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00604/full#supplementary-material>

REFERENCES

- Angioi, S. A., Rau, D., Attene, G., Nanni, L., Bellucci, E., Logozzo, G., et al. (2010). Beans in Europe: origin and structure of the European landraces of *Phaseolus vulgaris* L. *Theor. Appl. Genet.* 121, 829–843. doi: 10.1007/s00122-010-1353-2
- Asfaw, A., Blair, M. W., and Almekinders, C. (2009). Genetic diversity and population structure of common bean (*Phaseolus vulgaris* L.) landraces from the East African highlands. *Theor. Appl. Genet.* 120, 1–12. doi: 10.1007/s00122-009-1154-7
- Bellucci, E., Nanni, L., Biagetti, E., Bitocchi, E., Giardini, A., Rau, D., et al. (2014). “Common bean origin, evolution and spread from America,” in *Phaseolus: A New World gift to mankind Why common beans are so common?*, ed D. Rubiales (Córdoba, Spain: International Legume Society), 2, 12–16.
- Bitocchi, E., Bellucci, E., Giardini, A., Rau, D., Rodriguez, M., Biagetti, E., et al. (2013). Molecular analysis of the parallel domestication of the common bean in Mesoamerica and the Andes. *New Phytol.* 197, 300–331. doi: 10.1111/j.1469-8137.2012.04377.x
- Bitocchi, E., Nannia, L., Belluccia, E., Rossia, M., Giardinia, A., Spagnoletti Zeulib, P., et al. (2012). Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U.S.A.* 109, 788–796. doi: 10.1073/pnas.1108973109
- Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., and Cavalli-Sforza, L. L. (1994). High resolution human evolutionary trees with polymorphic microsatellites. *Nature* 368, 455–457. doi: 10.1038/368455a0
- Burger, J. C., Chapman, M. A., and Burke, J. M. (2008). Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95, 113–122. doi: 10.3732/ajb.95.2.113
- Cardador-Martínez, A., Loarca-Piña, G., and Oomah, B. D. (2002). Antioxidant activity in common beans (*Phaseolus vulgaris* L.). *J. Agric. Food. Chem.* 50, 6975–6980. doi: 10.1021/jf020296n
- Chávez-Servia, J. L., Heredia-García, E., Mayek-Pérez, N., Aquino-Bola-os, E. N., Hernández-Delgado, S., Carrillo-Rodríguez, J. C., et al. (2016). “Diversity of common bean (*Phaseolus vulgaris* L.) landraces and the nutritional value of their grains,” in *Grain Legumes*, ed A. Goyal (Rijeka: InTech Open Access Publisher), 1–33.
- Čupić, T., Gantner, R., Popović, S., Tucak, M., and Sudar, R., Stjepanović, M. (2012). “Rasprostranjenost jednogodišnjih mahunarki u Hrvatskoj,” in *Proceedings & Abstracts of the 5th International Scientific/Professional Conference Agriculture in Nature and Environment Protection*, eds B. Stipešević and R. Sorić (Osijek: Glas Slavonije d.d.), 220–225.
- De la Fuente, M., López-Pedrou, M., Alonso, J., Santalla, M., De Ron, A. M., Álvarez, G., et al. (2012). In-depth characterization of the Phaseolin protein diversity of common bean (*Phaseolus vulgaris* L.) based on twodimensional electrophoresis and mass spectrometry. *Food Technol. Biotechnol.* 50, 315–325.
- Earl, D. A., and von Holdt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Res.* 4, 359–361. doi: 10.1007/s12686-011-9548-7
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Excoffier, L., Laval, G., and Schneider, S. (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol. Bioinform.* 1, 47–50.
- Excoffier, L., Smouse, P. E., and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131, 479–491.
- FAO (2008). *Country Report on the State of Plant Genetic Resources for Food and Agriculture, Croatia*. Available Online at: <http://www.fao.org/docrep/013/i1500e/Croatia.pdf>
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678
- Felsenstein, J. (2004). *PHYMLIP (Phylogeny Inference Package) Version 3.6*. Department of Genomic Sciences, University of Washington, Seattle, WA.
- Gepts, P., and Debouck, D. G. (1991). “Origin, domestication and evolution of the common bean, *Phaseolus vulgaris*,” in *Common Beans: Research for Crop Improvement*, eds O. Voysey and A. Van Schoonhoven, (Oxon, UK: CAB), 7–53.
- Gioia, T., Logozzo, G., Attene, G., Bellucci, E., Benedettelli, S., Negri, V., et al. (2013). Evidence for Introduction Bottleneck and Extensive Inter- Gene Pool (Mesoamerica x Andes) Hybridization in the European Common Bean (*Phaseolus vulgaris* L.) Germplasm. *PLoS ONE* 8:e75974. doi: 10.1371/journal.pone.0075974
- Goudet, J. (2002). *FSTAT. A Program for Windows to Estimate and Test Gene Diversities and Fixation Indices. Version 2.9.3*. Available online at: <http://www2.unil.ch/popgen/softwares/fstat.htm>
- Hammer, K. (1984). Das Domestikationssyndrom. *Kulturpflanze* 32, 11–34. doi: 10.1007/BF02098682
- Kami, J., Becerra Velasquez, V., Debouck, D. G., and Gepts, P. (1995). Identification of presumed ancestral DNA sequences of phaseolin in *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. U.S.A.* 92, 1101–1104. doi: 10.1073/pnas.92.4.1101
- Koinange, E. M. K., Singh, S. P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci.1996.0011183X003600040037x
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., and Mayrose, I. (2015). Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Res.* 15, 1179–1191. doi: 10.1111/1755-0998.12387
- Kwak, M., and Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theor. Appl. Genet.* 118, 979–992. doi: 10.1007/s00122-008-0955-4
- Kwak, M., Toro, O., Debouck, D. G., and Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Ann. Bot.* 110, 1573–1580. doi: 10.1093/aob/mcs207

- Lioi, L. (1989). Geographical variation of phaseolin patterns in an old world collection of *Phaseolus vulgaris*. *Seed Sci. Technol.* 17, 317–324.
- Lioi, L., Piergiovanni, A. R., Pignone, D., Puglisi, S., Santantonio, M., and Sonnante, G. (2005). Genetic diversity of some surviving on-farm Italian common bean (*Phaseolus vulgaris* L.) landraces. *Plant Breed.* 124, 576–581. doi: 10.1111/j.1439-0523.2005.01153.x
- Lithourgidis, A. S., Dordas, C. A., Damalas, C. A., and Vlachostergios, D. N. (2011). Annual intercrops: an alternative pathway for sustainable agriculture. *Aust. J. Crop Sci.* 5, 396–410.
- Logozzo, G., Donnoli, R., Macaluso, L., Papa, R., Knupffer, H., and Spagnoletti Zeuli, P. L. (2007). Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. *Genet. Res. Crop Evol.* 54, 1763–1779. doi: 10.1007/s10722-006-9185-2
- Maras, M., Pipan, B., Šuštar-Vozlič, J., Todorović, V., Šurić, G., Vasić, M., et al. (2015). Examination of genetic diversity of common bean from the Western Balkans. *J. Amer. Soc. Hort. Sci.* 140, 308–316.
- Matsuoka, Y., Vigouroux, Y., Goodman, M. M., Sanchez, J. G., Buckler, E., and Doebley, J. (2002). A single domestication for maize shown by multilocus microsatellite genotyping. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6080–6084. doi: 10.1073/pnas.052125199
- Minch, E., Ruiz-Linares, A., Goldstein, D., Feldman, M., and Cavalli-Sforza, L. L. (1997). *MICROSAT: a Computer Program for Calculating Various Statistics on Microsatellite Allele Data, Version 1.5d*. Stanford, CA: Stanford University.
- Papa, R., Nanni, L., Sicard, D., Rau, D., and Attene, G. (2006). “The evolution of genetic diversity in *Phaseolus vulgaris* L.,” in *Darwin’s Harvest: New Approaches to the Origins, Evolution and Conservation of Crops*, eds T. J. Motley, N. Zerega, and H. Cross (New York, NY: Columbia University Press), 121–142. doi: 10.7312/motl13316-007
- Pathania, A., Kumar Sharma, S., and Nath Sharma, P. (2014). “Common Bean,” in *Broadening the Genetic Base of Grain Legumes*, eds M. Singh, I. Singh Bisht, and M. Dutta, (New Delhi: Springer), 11–50.
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*. 155, 945–959.
- Radosavljević, I., Jakse, J., Javornik, B., Satovic, Z., and Liber, Z. (2011). New microsatellite markers for *Salvia officinalis* (Lamiaceae) and cross-amplification in closely related species. *Am. J. Bot.* 98, 316–318. doi: 10.3732/ajb.1000462
- Raggi, L., Tiranti, B., and Negri, V. (2013). Italian common bean landraces: diversity and population structure. *Genet. Res. Crop Evol.* 60, 1515–1530. doi: 10.1007/s10722-012-9939-y
- Raymond, M., and Rousset, F. (1995). GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered.* 86, 248–249. doi: 10.1093/oxfordjournals.jhered.a111573
- Reynoso-Camacho, R., Ramos-Gomez, M., and Loarca-Pina, G. (2006). “Bioactive components in common beans (*Phaseolus vulgaris* L.),” in *Advances in Agricultural and Food Biotechnology*, eds R. Guevara-González and I. Torres-Pacheco (Trivandrum: Research Signpost), 217–236.
- Santalla, M., Rodiño, P., and De Ron, A. M. (2002). Allozyme evidence supporting southwestern Europe as a secondary center of genetic diversity for the common bean. *Theor. Appl. Genet.* 104, 934–944. doi: 10.1007/s00122-001-0844-6
- SAS Institute (2004). *SAS/STAT® 9.1 User’s Guide*. Cary, NC: SAS Institute Inc.
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Gen.* 46, 707–713. doi: 10.1038/ng.3008
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18, 233–234. doi: 10.1038/72708
- Singh, S. P. (1982). A key for identification of different growth habits of *Phaseolus vulgaris* L. *Ann. Rep. Bean Improv. Coop.* 25, 92–95.
- Svetleva, D., Pereira, G., Carlier, J., Cabrita, L., Leitao, J., and Genchev, D. (2006). Molecular characterization of *Phaseolus vulgaris* L. genotypes included in Bulgarian collection by ISSR and AFLP analyses. *Sci. Hort.* 109, 198–206. doi: 10.1016/j.scienta.2006.04.001
- Zizumbo-Villarreal, D., and Colunga-García, P. (2010). Origin of agriculture and plant domestication in West Mesoamerica. *Genet. Res. Crop Evol.* 57, 813–825. doi: 10.1007/s10722-009-9521-4

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Carović-Stanko, Liber, Vidak, Barešić, Grdiša, Lazarević and Šatović. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Reproductive Strategies in Mediterranean Legumes: Trade-Offs between Phenology, Seed Size and Vigor within and between Wild and Domesticated *Lupinus* Species Collected along Aridity Gradients

Jens D. Berger^{1,2*}, Damber Shrestha² and Christiane Ludwig¹

¹ CSIRO Agriculture and Food, Wembley WA, Australia, ² Centre for Legumes in Mediterranean Agriculture, Faculty of Natural and Agricultural Sciences, University of Western Australia, Crawley, WA, Australia

OPEN ACCESS

Edited by:

Diego Rubiales,
Instituto de Agricultura Sostenible
(CSIC), Spain

Reviewed by:

Fred Stoddard,
University of Helsinki, Finland
Rafael Rubio de Casas,
University of Granada, Spain

*Correspondence:

Jens D. Berger
Jens.Berger@csiro.au

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 23 December 2016

Accepted: 27 March 2017

Published: 13 April 2017

Citation:

Berger JD, Shrestha D and Ludwig C
(2017) Reproductive Strategies in
Mediterranean Legumes: Trade-Offs
between Phenology, Seed Size and
Vigor within and between Wild and
Domesticated *Lupinus* Species
Collected along Aridity Gradients.
Front. Plant Sci. 8:548.
doi: 10.3389/fpls.2017.00548

To investigate wild and domesticated Mediterranean annual reproductive strategies, common garden comparisons of Old World lupins collected along aridity gradients were initiated. These are excellent candidates for ecophysiology, being widely distributed across contrasting environments, having distinct domestication histories, from ancient *Lupinus albus* to recently domesticated *Lupinus angustifolius* and *Lupinus luteus*, facilitating the study of both natural and human selection. Strong trade-offs between seed size, early vigor and phenology were observed: vigor increasing, and flowering becoming earlier with increasing seed size. Despite large specific differences in all these traits, natural and human selection have operated in very similar ways in all 3 species. In wild material, as collection environments became drier and hotter, phenology became earlier, while seed size, early vigor and reproductive investment increased. Wild and domesticated germplasm separated along similar lines. Within similar habitats, domesticated material was consistently earlier, with larger seeds, greater early vigor and higher reproductive investment than wild, suggesting selection for both early establishment and timely maturity/drought escape in both domesticated and wild low rainfall ecotypes. Species differences reflected their distribution. Small and soft-seeded, low vigor *L. luteus* had a late, rainfall-responsive phenology specifically adapted to long season environments, and a narrow coastal distribution. *L. angustifolius* was much more conservative; more hard-seeded, flowering and maturing much earlier, with a wide Mediterranean distribution. *L. albus* flowered earlier but matured much later, with longer reproductive phases supporting much larger seed sizes and early vigor than either *L. luteus* or *L. angustifolius*. This ruderal/competitive combination appears to give *L. albus* a broad adaptive capacity, reflected in its relatively wider Mediterranean/North African distribution.

Keywords: adaptation, crop evolution, terminal drought, phenology, seed size, early vigor, hard seed breakdown

INTRODUCTION

Despite their high value and rotational benefits, cool season legume crops can struggle to retain their place in farming systems because of their perceived riskiness (Pannell, 1995). There is a wide range of biotic (weeds, pests and diseases) and abiotic stresses, such as drought, heat and cold that impact grain legume production. Understanding species' adaptive responses to such stresses is the key to genetic risk mitigation, by ensuring that crops with an appropriate assemblage of adaptive traits are grown in fitting production niches. To define species adaptive potential it is important to work with a broad range of genetically diverse material, rather than the narrow band of elite cultivars that typifies many modern grain legume crops (Abbo et al., 2003; Berger et al., 2013). Genetic resource collections that sample a wide range of habitats are an ideal vehicle for this.

Here we focus on adaptive strategies among the “Old World” lupins, (*Lupinus albus* L., *L. angustifolius* L., *L. luteus* L.) widely collected along terminal drought stress gradients across the Mediterranean basin (Berger et al., 2008). The Mediterranean climate poses wide ranging challenges to plant adaptation. Despite the almost ubiquitous cool, wet winters and hot, dry summers there is wide spatial and temporal variation. Seasonal rainfall, temperature, relative humidity, solar radiation, and wind speed vary across the Mediterranean basin (Hijmans et al., 2005), coupled with differences in aspect, slope, soil depth, pH, fertility, water holding capacity, land management and grazing intensity. As a result of this manifold variation, broadly distributed species typically encounter contrasting habitats. These exert differential selection pressure, leading to the formation of distinct specifically adapted ecotypes. Understanding the extent to which the “Old World” lupins form distinct ecotypes will help breeders to maximize productivity in their target environments, and address gaps—such as the current lack of a long season narrow-leaved lupin cultivar (Berger et al., 2012a; Chen et al., 2016).

A range of reproductive strategies has been invoked to explain specific adaptation in Mediterranean winter annuals, typically emphasizing phenology and its trade-offs against other plant attributes. Grime's (1979) triangle balances phenology against biomass production/resource acquisition along the disturbance/competition continuum. In the Mediterranean context, Grime's (1979) triangle predicts that conservative ruderal strategies (rapid growth rates/short life cycles) are favored in environments “disturbed” by terminal drought: high temperatures, and low, uncertain rainfall. Conversely, cooler, longer season environments with high, consistent rainfall are more likely to select for longer lifecycles supporting the development of competitive traits that facilitate resource capture, such as high investment in above and below ground biomass, ultimately leading to greater seed production. Grime's (1979) predictions are well supported among Mediterranean annuals, with a strong focus on pasture legumes (see references in Berger and Ludwig, 2014). This conservative-competitive trade-off has far reaching ramifications. In *L. luteus*, in addition to phenology/productivity trade-offs, rates of water-use, timing of stress onset and drought tolerance capacity are all traded-off among low and high rainfall ecotypes (Berger and Ludwig,

2014). Indeed, we suggest that the expression of drought escape, postponement and tolerance is closely linked to the phenology-productivity trade-off, producing integrated adaptive strategies that suit specific environmental niches (Berger et al., 2016) that should be exploited by farmers.

Dormancy, dispersal and seed size are also traded off in adaptation. Dormancy and dispersal facilitate escape from stress in space and time, while larger seeds increase the probability of successful establishment under unfavorable conditions (Venable and Brown, 1988). Accordingly, seed size is predicted to increase as dormancy or dispersal decrease (Venable and Brown, 1988), while the latter are also negatively correlated, since having more of one reduces the need for the other (Rubio de Casas et al., 2015). Germination behavior has been divided into 2 categories (Gremer et al., 2016). Predictive or plastic germination that responds to cues signaling the onset of favorable conditions for seedling growth are expected to be important in predictable climates, and habitats where strong competitive pressure selects for early emergence to acquire resources before they are lost. Alternatively, bet-hedging germination, where cohorts of seed germinate at different times spreads risk in time. Diversified bet-hedging trades off reduced fitness at any given time point (because only a fraction of the population is able to germinate) against reduced variance in fitness over time, and is only adaptive when environmental conditions vary from generation to the next (Donaldson-Matasci et al., 2013).

In Mediterranean annual legumes physical dormancy (also referred to as hardseededness), caused by a water-impermeable seed coat that tends to break down during fluctuating summer temperatures, is the most important mechanism preventing inopportune germination (Norman et al., 2002), protecting populations against false breaks before the onset of the rainy season. If there is variability in the degree of physical dormancy loss this mechanism may also play a diversified bet-hedging role. Therefore, given that rainfall variability tends to increase with aridity, hard seed incidence may increase in low rainfall environments (Hacker, 1984; Hacker and Ratcliff, 1989; Piano et al., 1996). Alternatively, in-season phenology may itself be traded-off against hardseededness (Norman et al., 2005). Genotypes with a high proportion of hard seed and low rate of dormancy loss can afford a more competitive, later phenology than those with soft, permeable seeds, if their additional productivity during favorable years establishes a larger seedbank than those of earlier, more ruderal genotypes. Phenology may also be traded-off against seed size because smaller seeds fill more quickly, allowing small seeded species to run a later phenology than large seeded species from the same environment (e.g., small seeded *Trifolium clusii* and *T. glanduliferum* vs. large seeded *T. purpureum* in Northern Israel) (Cocks, 1999). Finally, increasing seed size and its associated competitive advantages are traded off against fecundity (Venable and Brown, 1988; see also references in Norman et al., 2005).

Clearly there are a wide range of strategies that allow annual plants to adapt to Mediterranean climates. While studies within and between widely distributed species are most informative (Thompson, 2005), the study of annual species has often focused on a narrow germplasm and geographic range (see references

in Berger and Ludwig, 2014). The “Old World” lupin species of the present study are interesting because they are large seeded, and adapted to neutral to acid, low water-holding capacity soils. Consequently seed size trade-offs are investigated at a different scale than the smaller seeded pasture legumes of previous studies (Cocks, 1999; Norman et al., 2005). Furthermore, phenology/productivity trade-offs may be stronger than in species from higher water holding capacity soils that buffer the onset of terminal drought to some extent. These include many of the species listed in Berger and Ludwig (2014), such as *Triticum dicoccoides* (Kato et al., 1998), *Hordeum spontaneum* (Volis, 2007), *Medicago* (Yousfi et al., 2010), and *Trifolium* species (Norman et al., 2002), typically collected from finer-textured sand- or clay-loams. Moreover, these “Old World” lupins have contrasting distributions, seed sizes, and domestication histories. *L. albus* is a very widely distributed Bronze Age domesticate with large to very large seeds grown on mildly alkaline to acidic sands and loams. Conversely, the smaller seeded *L. luteus* and *L. angustifolius*, domesticated in the last 200 years, have limited to intermediate natural Mediterranean distributions on acid, sandy soils (Gladstones, 1974; Hondelmann, 1984; Huyghe, 1997; Zohary and Hopf, 2000).

Given these differences, we were interested to discover the extent to which these Old World lupins trade-off phenology, productivity, seed size and hardseededness along collection site terminal drought stress gradients. Is there a single, common reproductive strategy within the genus for coping with the transition from low to high rainfall environments, or are there specific differences which relate to seed size, distribution or domestication history? Beyond the accepted

“domestication syndrome” traits (Meyer et al., 2012), how do wild and domesticated old world lupins differ, and did ancient and modern domestication select for similar or different reproductive strategies?

MATERIALS AND METHODS

Germplasm Selection and Habitat Characterization

All experiments in the present study were based on the evaluation of wild and domesticated old world lupins (*L. albus*, *L. angustifolius* and *L. luteus*) collected from a wide range of habitats representing each species’ distribution range (Figure 1, Table 1; Berger et al., 2008). Collection site climate was characterized in Berger et al. (2008) by defining typical growing season phenology for each site and calculating bioclimatic variables for vegetative and reproductive phases based on long term average precipitation (monthly totals and coefficients of variation), temperature, daylength, relative humidity, sunshine percentage, wind speed, and precipitation (Hijmans et al., 2005). Relationships among these bioclimatic variables, as well as latitude, longitude and altitude were simplified using principal components analysis (PCA) performed within each species. The resultant scores for PC 1–4, capturing 78.2–90.6% of variance, were used to classify collection sites by hierarchical clustering (Ward’s method, SPSS Version 10).

These Ward’s clusters formed the basis of the comparisons made in the present study. All 3 species were distributed along terminal drought stress gradients defined by reproductive phase rainfall and temperature; stress increasing with cluster

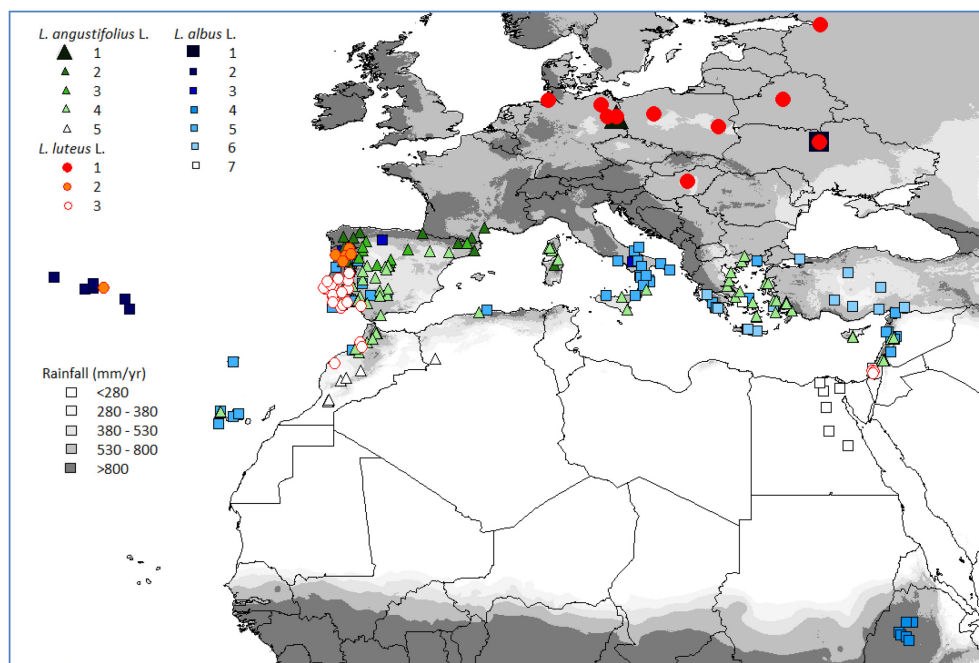


FIGURE 1 | Northern hemisphere collection sites of old world lupin genotypes evaluated in the present study, classified by species and habitat cluster (see Table 1).

TABLE 1 | Provenance and seasonal collection site climate (details in Berger et al., 2008) of old world lupin common garden field experiments at in 2007, 2008 and 2015 (hard seed breakdown).

aClust	Description	Germplasm n			Rainfall (mm)		Rain days		Frost days	Mean temp (°C)		Temp change (°C/day)	Summer temp (°C)		bStatus			Origin		
		07	08	15	Pre-seas	Veg	Rep	Veg		Rep	Max		Range	Br	Lr	W				
1	°L. albus L. European spring-sown: long daylength, rapidly warming, cool veg phase, med rainfall, low terminal drought.	88	17		310	102	226	24	42	7	12.0	18.9	0.01	24.7	11.4	14	58	16	UKR, 1	
		6	2		104	515	521	69	96	4	14.5	15.8	0.05	23.8	9.4		6			
		5			54	581	234	81	42	56	5.8	15.0	0.09	23.6	12.5		2	3		ITA, 3; ESP, 1; PRT, 1
		6	3		773	688	61	33	9	2	16.2	16.3	0.01	27.0	16.3		6			ETH, 6
5	Average Mediterranean climates: intermediate terminal drought.	48	6		41	376	234	47	41	18	12.2	16.3	0.09	29.2	13.3		36	12	ITA, 15; ESP, 11; PRT, 8; GRC, 6; SYR, 6; MAR, 1; DZA, 1	
6	Mediterranean spring-sown: low rainfall, warm rep phase with rapid temp increase; high terminal drought.	14	3		552	91	13	19	4	2	16.4	23.6	0.10	30.0	14.0		14		TUR, 7; GRC, 7	
7	Southern Mediterranean: low rainfall, warm rep phase with rapid temp increase; high terminal drought.	8	3		5	47	32	8	10	3	16.6	18.2	0.12	34.6	14.8		8		EGY, 6; ISR, 1; GRC, 1	
1	L. angustifolius L. European spring-sown: long daylength, rapidly warming, cool veg phase, med rainfall, low terminal drought.	135	42	21												16	0	115	DEU, 1	
		1	3		269	91	186	26	40	11	11.1	17.9	0.02	23.8	12.1		1			
2	Mediterranean long season: high, frequent rainfall, cool veg phase; low terminal drought.	18	8	8	64	558	324	76	62	50	8.3	15.6	0.09	26.4	13.8		18		ESP, 10; PRT, 4; FRA, 3; ITA, 1	

(Continued)

TABLE 1 | Continued

aClust	Description	Germplasm n			Rainfall (mm)		Rain days		Frost days		Mean temp (°C)		Temp change (°C/day)		Summer temp (°C)			bStatus			Origin
		07	08	15	Pre-seas	Veg	Rep	Veg	Rep	Veg	Rep	Rep	Max	Range	Br	Lr	W				
3	Mediterranean (Iberian) long season (cool): intermediate, frequent rainfall, cold frosty veg phase; med terminal drought.	13	1		50	391	164	85	39	97	5.7	16.0	0.10		26.8	16.7			11	ESP, 11	
4	Average Mediterranean climates: intermediate terminal drought.	92	23	9	46	345	226	44	38	16	12.2	15.9	0.10		31.2	15.6	15		77	ESP, 20; AUS, 19; GRC, 17 ; MAR, 10; ISR, 5; FRA, 4; ITA, 4; PRT, 3; SYR, 3; TUR, 3; CYP, 2; DZA, 1; ZAF, 1	
5	Nth Africa: low, infrequent rainfall, high sun hours, rainfall variability: high terminal drought.	7	5	4	26	145	135	18	20	21	11.4	16.3	0.10		33.7	19.6			7	MAR, 6; DZA, 1	
	Unclassified	4	2														2		2	GRC, 2; CHL, 2	
1	L. luteus L. European spring-sown: long daylength, rapidly warming, cool veg phase, med rainfall, low terminal drought.	71	29	13	296	99	211	26	41	11	10.9	17.7	0.01		23.6	11.7	35	0	36	DEU, 6; BYS, 5; UKR, 5; HUN, 4; POL, 3; AUS, 2; RUS, 2; SUN, 1; LVA, 1; NLD, 1	
2	Mediterranean (Iberian) long season: high, frequent rainfall, cool veg phase; low terminal drought.	7	7	6	70	645	416	76	72	27	10.5	15.5	0.08		26.2	13.2			7	PRT, 6; ESP, 1	
3	Average Mediterranean climates: intermediate terminal drought.	32	11	6	47	368	224	45	40	4	13.5	16.2	0.08		30.1	15.2	5		29	PRT, 17; AUS, 9; MAR, 3; ISR, 3; ESP, 1; ZAF, 1	
	Unclassified	2																		PRT, 2	
COMMON GARDEN FIELD TRIAL CLIMATE																					
	2007 field trial				230	262	69	46	21	0	14.1	17.1		0.11							
	2008 field trial				258	187	101	32	36	0	12.5	16.3		0.07							

COMMON GARDEN FIELD TRIAL CLIMATE

2007 field trial	230	262	69	46	21	0	14.1	17.1	0.11
2008 field trial	258	187	101	32	36	0	12.5	16.3	0.07

^aCluster: collection sites were clustered largely on the basis of seasonal climate and described in Berger et al. (2008).
^bStatus: Br, breeding material or released cultivar; Lr, landrace; W, wild germplasm. Numbers refer to 2007 evaluation.
^c*L. albus* L. collection site climates above were re-analysed in light of the fact that landraces from the Balkans to Asia Minor are spring- (Uluyim et al., 2002; Mihailovic et al., 2008), rather than autumn-sown as assumed in Berger et al. (2008).
Note that 2007 and 2008 field trial climate is also included to put these evaluations into context alongside germplasm collection site climate. Summer maxima and average temperature range are included as background for the physical dormancy work.

number (Table 1). Thus, Cluster 1 (spring-sown Central Europe) had the lowest terminal drought stress, increasing through a range of Mediterranean regions, typically culminating in rapidly warming, low rainfall reproductive phase northern African environments (e.g., *L. angustifolius*, Cluster 5; *L. albus*, Cluster 7). *L. albus* was particularly widely distributed, from the Azores and Canary islands, through the Mediterranean basin, including the Nile valley and Ethiopian highlands (Figure 1). *L. albus* was largely domesticated (Table 1), with many Mediterranean, but also Ethiopian landraces (Cluster 3), reflecting its long history as a domesticated crop relative to other Old World lupins (Zohary and Hopf, 2000). Conversely, *L. luteus* and especially *L. angustifolius* were dominated by wild Mediterranean germplasm, but also included some cultivars from Europe and Mediterranean climates in Western Australia and South Africa. *L. luteus* was distributed less widely than the others, along a narrower terminal drought stress gradient, because even the most stressful habitat (Cluster 3) has relatively high reproductive phase rainfall and only a moderate temperature increase compared to *L. angustifolius* and *L. albus* (Table 1).

2007 Field Evaluation

Lupinus albus ($n = 88$), *Lupinus angustifolius* ($n = 133$), and *Lupinus luteus* ($n = 73$) were acquired from the Australian Lupin Collection and evaluated as spaced plants in a common garden field experiment (RCBD, $n = 3$) in a sandy loam at CSIRO Floreat (-31.95 N, 115.79 E). Plots were hand planted on 9th July 2007 after vernalization to simulate an endemic Mediterranean environment, where lupins are vernalized by low temperatures early in their lifecycle. Seeds were scarified and allowed to swell at room temperature overnight, and then placed into moist petri dishes in dark growth cabinets for vernalization at 8°C for 16 days, using P-Pickle-T[®] prophylaxis against fungal infection. Seeds were inoculated with Nodulaid 100[®] (Group G rhizobia) immediately prior to planting.

Early vigor was estimated by measuring plant biomass and height at 625°d (45 days after sowing, assuming base temperature = 0°C). To estimate growth rates non-destructively, plant height was measured on 2 subsequent occasions (1023 , 1389°d ; 71, 95 days) during the linear growth phase. Phenological observations (onset and end of flowering, onset of podding, maturity) were made 3 times weekly from flowering onwards. At maturity (defined by 95% of pods being dry) all above-ground biomass was harvested, separated into vegetative and reproductive matter by branch order (mainstem, lateral and basal), and weighed after oven drying at 60°C for 48 h. Seed and pod numbers were counted and weighed within each branch order to facilitate the calculation of separate yield components (seed size, seeds per pod).

2007 growing season rainfall was low (331 mm), and the trial exposed to terminal drought, with little reproductive phase rainfall combined with relatively high temperatures and a high rate of temperature increase (Table 1).

2008 Field Evaluation

To validate the 2007 results, a smaller, representative subset of wild and domesticated germplasm (*L. albus*, $n = 17$; *L. angustifolius*, $n = 42$; *L. luteus*, $n = 29$) was evaluated in a common garden in deep sands at the adjacent University of Western Australia field station (-31.95 N, 115.79 E) in 2008 using seed from the Australian Lupin Collection. Seeds were imbibed, vernalized at 4°C for 28 days prior to planting and hand planted in a RCBD ($n = 4$) on 19th June 2008 in single row plots (1.5 m long, 10 cm inter-plant, 20 cm inter-row distances). Phenological observations (flowering, end of flowering, maturity) were made 3 times weekly from flowering onwards. Biomass was destructively harvested in 1–3 plants per plot, and canopy height measured non-destructively in 5 adjacent plants per plot during the linear growth phase at 604, 905, 1450, and 1850°d (corresponding to 48, 72, 110 and 131 days).

While total growing season rainfall was lower in 2008 than 2007 (288 vs. 331 mm), reproductive phase rainfall was higher, considerably more frequent, and coincided with lower temperatures than in 2007 (Table 1).

Loss of Seed Dormancy

To investigate the role of seed dormancy loss in adaptation to rainfall gradients, low and high rainfall subsets of wild *L. angustifolius* ($n = 20$) and *L. luteus* ($n = 12$) were selected. This rainfall contrast also has implications on summer temperatures. Both the summer maxima and diurnal temperature range tend to increase with aridity, rising with cluster number in these species (Table 1). Accordingly, the contrasting rainfall subsets of wild *L. angustifolius* and *L. luteus* investigated for loss of physical dormancy also represent contrasting summer temperatures (Table 2). Seed dormancy loss was not investigated in *L. albus* because in this species the distinction between wild material and escaped landraces is tenuous.

Seeds were multiplied in a common field at CSIRO Floreat (-31.95 N, 115.79 E) in 2014, harvested in November and immediately returned to the field to simulate Mediterranean climate hard seed breakdown in diurnally varying summer temperatures from December onwards according to the following protocol. 50 seeds per replication (RCBD, $n = 4$) were bagged in mesh pouches, and placed on the soil surface of the

TABLE 2 | Average seasonal rainfall and summer temperatures (mean, maximum and diurnal range) in high and low rainfall subsets of wild *L. angustifolius* ($n = 20$) and *L. luteus* ($n = 12$) used in the investigation of hard seed breakdown over time.

Species/Category	Season rain (mm)	Summer temperatures ($^{\circ}\text{C}$)		
		Mean	Maximum	Diurnal range
<i>L. angustifolius</i> L.	718	22.0	30.0	16.5
High rainfall	1022	20.7	28.2	15.5
Low rainfall	262	24.0	32.7	18.0
<i>L. luteus</i> L.	773	22.6	28.8	13.4
High rainfall	1064	19.7	25.7	12.7
Low rainfall	482	25.5	31.9	14.1

CSIRO rainout shelter amongst retained crop stubble on 18th December 2014. At approximately monthly intervals the seeds were allowed to imbibe water for 48 h, after which swollen and germinating seeds were recorded and removed. Remaining hard seeds were returned to the mesh pouches in the field and the process repeated to generate cumulative hard seed breakdown curves. Note that between December and May the rainout shelter was used to avoid uncontrolled seed wetting, whereas during the May–December growing season the pouches were exposed to rain as well as monthly imbibition for 48 h.

Statistical Analysis

The data was analyzed separately for each year using Genstat V16. Within and between species differences were investigated using ANOVA and linear regression, fitting species as main effects, clusters (**Table 1**) nested within species, and accessions nested within clusters. In all analyses, residual plots were generated to identify outliers, and confirm that variance was common and normally distributed. Transformations were made as appropriate. Principal components analysis (PCA) was used to describe the relationships between collection site bioclimatic data and within- and between-species biology using data in Berger et al. (2008) and ANOVA means from the 2007 and 2008 field experiments. PCA was based on a correlation matrix to remove scale effects between variables.

RESULTS

Phenology and Environment

To demonstrate the interaction between habitat climate and lupin phenology, PCA was performed on accession collection site bioclimatic and 2007 field phenology data in a single analysis. If phenology-climate matching is a key adaptation strategy in lupin, we expect to find vectors describing accession phenology and site climate aligned on the same dimension in the PCA. PC1 captured a drought stress gradient: seasonal sunshine hours, rainfall variability and vegetative phase temperature increasing; and latitude, seasonal relative humidity, rainfall frequency and reproductive phase rainfall decreasing from right to left on PC1, in inverse proportion to the aforementioned drought stress gradient, particularly those vectors describing reproductive rainfall, rainfall frequency and average vegetative phase temperature. PC2 was dominated by negative loadings on vegetative phase daylength, pre-season rainfall, and positive loadings on maximum seasonal temperature and rates of reproductive phase temperature increase; separating the cooler domesticated European (all 3 species) and Ethiopian highland *L. albus* collection sites from their Mediterranean counterparts.

All 3 lupin species were broadly distributed along PC1 (particularly *L. albus*), indicative of a common, wide ranging stress gradient among collection sites, and a correspondingly common, wide phenological response. To investigate this directly, flowering date was regressed against collection site seasonal rainfall (**Figure 2B**). While flowering date was delayed

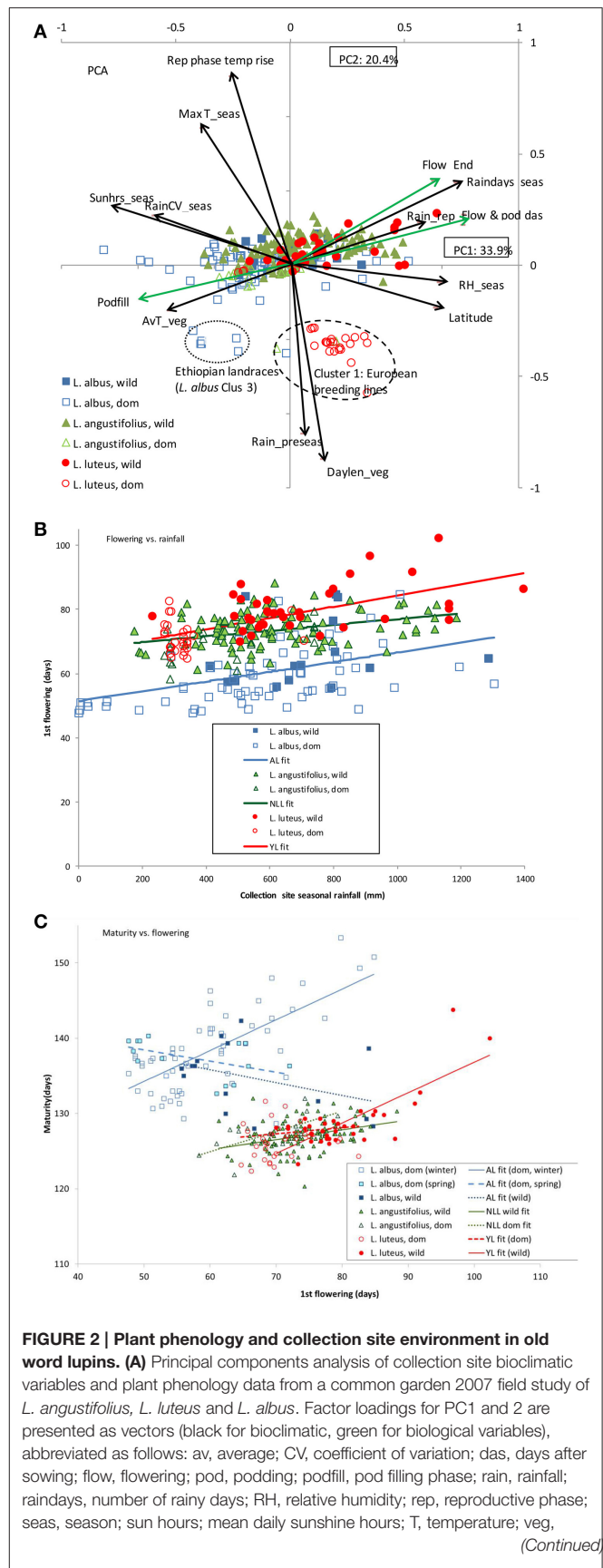


FIGURE 2 | Plant phenology and collection site environment in old world lupins. (A) Principal components analysis of collection site bioclimatic variables and plant phenology data from a common garden 2007 field study of *L. angustifolius*, *L. luteus* and *L. albus*. Factor loadings for PC1 and 2 are presented as vectors (black for bioclimatic, green for biological variables), abbreviated as follows: av, average; CV, coefficient of variation; das, days after sowing; flow, flowering; pod, podding; podfill, pod filling phase; rain, rainfall; raindays, number of rainy days; RH, relative humidity; rep, reproductive phase; seas, season; sun hrs, mean daily sunshine hours; T, temperature; veg, (Continued)

FIGURE 2 | Continued

vegetative phase. Markers represent genotype scores, classified by species and domestication status. **(B)** Linear regression of flowering against collection site season rainfall, fitting species as factors, accounting for 55.8% of variance. Regression equations presented in **Table 5**. **(C)** Linear regression of maturity against flowering, fitting species and domestication status as factors, accounting for 79.5% of variance. Regression equations presented in **Table 5**.

as rainfall increased in all 3 species ($P < 0.001$), there were significant intercept and slope differences. *L. luteus* and *L. albus* were similarly responsive to collection site seasonal rainfall (16.4 days delay/1000 ml rainfall, $P_{\text{diff}} = 0.538$), but *L. luteus* flowered consistently 15 days later than *L. albus* across its rainfall range (**Figure 2B**, $P < 0.001$). *L. angustifolius* and *L. luteus* intercepts were similar (67 days, $P_{\text{diff}} = 0.607$), but *L. angustifolius* flowering was much less responsive (9 days delay/1000 ml rainfall, $P_{\text{diff}} = 0.03$), causing these species to diverge under high collection site rainfall (**Table 3**). Thus, in average Mediterranean climates, wild *L. albus* flowered very early (Cluster 5, 65 days), while *L. angustifolius* (Cluster 4, 73 days) was only marginally earlier than *L. luteus* (Cluster 3, 78 days). Conversely, in cool, high rainfall, long-season Iberian habitats, *L. albus* flowered in 70 days (Cluster 2), followed by *L. angustifolius* (Cluster 3, 78 days), while *L. luteus* was much later (Cluster 2, 89 days).

While flowering and maturity date tended to be positively correlated, there were again considerable differences within and between species (**Figure 2C**, **Table 3**). *L. albus* tended to mature much later than the other species. While, domesticated, winter-sown *L. albus* had a very late, strongly flowering-responsive maturity, its spring-sown and wild counterparts did not delay maturity as flowering date increased (**Figure 2C**). *L. angustifolius* matured earliest (**Table 3**), was less flowering-responsive than the remaining species, and there were no differences between domesticated and wild material (**Figure 2C**, $P_{\text{diff}} = 0.278$), nor between clusters (**Table 3**). Domesticated *L. luteus* was similar to *L. angustifolius*, while wild material was as flowering-responsive as domesticated, winter-sown *L. albus* (**Figure 2C**), driven by late flowering and maturity in Cluster 2 (**Table 3**).

Phenology, Growth, and Productivity within and between Species

To highlight the effect of this phenological variability on growth and productivity, a combined species PCA was performed, accounting for 64% of variance in 2 components (**Figure 3**). PC1 captured a phenology-productivity continuum with negative loadings on flowering and podding, and positive loadings on seed size, early vigor, length of the reproductive phase (pod fill), harvest index, plant growth rates, height and productivity. PC2 was dominated by positive loadings on seed number and vegetative biomass.

Despite considerable within-species variation, the 3 old world lupin species plotted discretely on **Figure 3**, reflecting considerable between species variance for all traits in the 2007 field evaluation (**Table 3**). *L. luteus* was located on the left of **Figure 3A**, indicative of late phenology, small seed size, low harvest index, growth rates and productivity (**Table 3**). Within

L. luteus there were clear cluster differences in both PC1 and 2 (**Figure 3A**). Wild germplasm, particularly Cluster 3, tended to have more positive PC2 scores than domesticated material, associated with higher seed numbers per plant (**Table 3**). Domesticated PC1 scores were more positive than wild, reflecting strong selection for early phenology, high early vigor, harvest index and larger seed size (**Table 3**: Cluster 1, Cluster 3, Br vs W). A similar contrast occurred within wild *L. luteus*: with higher PC1 scores, earlier flowering, higher vigor, and productivity in low than high rainfall ecotypes (**Table 3**: Clusters 3, W vs. 2, W).

Despite a wide PC1 range, most *L. albus* accessions were located in the lower-right quadrant of **Figure 3B**, characterized by early phenology, large seeds, high early vigor, rapid growth rates, and high plant height, harvest index, seed and biological yield (**Table 3**). PC2 scores were predominantly negative in *L. albus*, indicative of low seed number. There were pronounced cluster differences along the phenology-productivity continuum in PC1, from the very early warm season, southern and spring-sown Mediterranean landraces (Clusters 7, 6) on the far right, through to the much later cool climate Iberian and Mediterranean germplasm (Clusters 2, 3) on the left. Cluster 5, representing average Mediterranean climates, was characterized by very wide ranging PC1 scores, exacerbated by the inclusion of 2 very late wild Greek outliers on the far left of PC1. Ethiopian *L. albus* (Cluster 4) formed a tightly clustered group in the middle of the phenology-productivity continuum. Wild-domesticated comparisons were feasible in Clusters 3 and 5. In both cases, PC1 scores were higher in domesticated than wild material, indicating similar selection for early phenology, high early vigor, harvest index and larger seed size as in *L. luteus* (**Table 3**).

Lupinus angustifolius was intermediate between *L. luteus* and *L. albus* along the PC1 phenology-productivity continuum, while PC2 scores tended to be more positive (**Figure 3C**), reflecting its greater fecundity (**Table 3**). Unlike the other 2 lupin species, *L. angustifolius* clusters were better aligned along a phenology-seed size/early vigor continuum described by the vectors mapping to the upper left and lower right quadrants of **Figure 3C**. Among the wild material, the cool climate Mediterranean and Iberian germplasm (Clusters 2, 3) was largely distributed along the upper left of **Figure 3C**, indicative of late phenology, small seed size and low early vigor (**Table 3**). Conversely, low rainfall North African material (Cluster 5) was located on the lower right, and characterized by early flowering, greater early vigor, and larger seed. Average Mediterranean climate (Cluster 4) germplasm was intermediate between these groups and very widely distributed, similar to its *L. albus* counterparts. Domesticated *L. angustifolius* was located on the lower-right extreme of **Figure 3C**, reflecting selection for earliness, high early vigor and large seed size (**Table 3**; Cluster 4, Br vs W). All *L. angustifolius* clusters ranged widely along a 45° axis defined by the lower left and upper right quadrants of **Figure 3C**, indicative of considerable within-cluster variation for plant height, growth rate and productivity.

Combined species PCA was repeated on the 2008 data with remarkably similar results, returning clear between and within species cluster separation along a phenology/vigor-drought stress continuum on PC1, and separating Mediterranean from domesticated European and Ethiopian material along

TABLE 3 | Within and between old world lupin species differences in terms of variance distribution^a and mean values for seed size, early growth, phenology and productivity in a rainfed field common garden evaluation in 2007 (See Table 4 for the 2008 common garden evaluation).

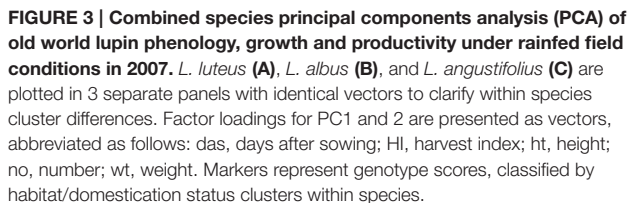
Category		Seed size	Early vigor	Growth rate	Plant ht	Flower days	Pod days	Rep phase	Maturity	HI	Biomass	Seed no/pl	Seed wt/pl (log)
^a Species variance		98***	96***	75***	89***	92***	92***	98***	97***	94***	84***	92***	92***
^a Cluster within species v.		2***	3***	19***	7***	7***	7***	2***	2***	4***	10***	6***	6***
^a Accession within cluster		0.2***	1***	6***	3***	1***	1***	0.3***	1***	2***	5***	2***	2***
Species means	<i>n</i>												
<i>L. albus</i> L.	84	26.8	0.68	0.93	57.8	61	73	64	138	38.4	69.8	107.4	1.3
<i>L. angustifolius</i> L.	127	10.1	0.29	0.80	48.5	73	81	46	127	34.8	56.9	201.9	1.2
<i>L. luteus</i> L.	69	11.9	0.34	0.86	50.6	76	84	43	128	30.4	53.5	137.8	1.1
LSD Species mean		1.2	0.05	0.08	1.3	1	1	1	1.2	3.0	9.7	15.2	0.1
Cluster means													
<i>L. albus</i> L.													
Albus 1, Br	1	21.5	0.26	0.41	32.3	56	67	67	129	30.7	16.3	23	0.6
Albus 2, LR	6	27.8	0.83	0.98	60.1	70	82	62	143	33.1	66.4	81	1.3
Albus 3, LR	2	23.5	0.71	1.06	64.6	67	77	62	143	46.1	61.0	119	1.3
Albus 3, W	3	18.6	0.50	0.74	45.3	72	82	50	133	41.3	45.4	93	1.2
Albus 4, LR	6	21.1	0.50	0.79	50.6	65	80	63	144	33.3	44.8	72	1.1
Albus 5, LR	42	30.2	0.77	0.98	60.9	59	71	66	137	39.0	75.4	103	1.4
Albus 5, W	12	23.8	0.52	0.88	52.8	65	76	59	135	37.4	64.4	109	1.3
Albus 6, LR	14	24.6	0.62	0.94	57.9	57	71	67	137	39.2	74.4	134	1.4
Albus 7, LR	8	30.1	0.75	0.92	61.2	52	65	70	135	45.7	84.4	140	1.5
LSD <i>albus</i>		2.8	0.25	0.25	9.0	4	4	5	4	10.0	29.3	57	0.3
<i>L. angustifolius</i> L.													
Angus 1, Br	1	14.5	0.43	0.80	50.5	74	80	47	127	40.2	36.0	101	1.2
Angus 2, W	17	7.3	0.24	0.91	52.1	80	87	41	128	30.6	47.7	213	1.0
Angus 3, W	11	8.2	0.25	0.88	51.0	78	85	42	127	34.6	52.1	218	1.1
Angus 4, Br	15	12.3	0.34	0.73	48.0	68	76	51	127	32.7	48.1	130	1.1
Angus 4, W	76	10.3	0.28	0.79	47.7	73	80	47	127	35.9	58.5	208	1.2
Angus 5, W	7	11.3	0.43	0.72	44.6	69	78	52	127	36.8	63.9	250	1.3
LSD <i>angustifolius</i>		2.1	0.13	0.23	8.5	4	4	5	3	9.9	34.9	130	0.2
<i>L. luteus</i> L.													
Luteus 1, Br	28	12.7	0.39	0.85	51.5	71	80	46	127	33.3	46.5	121	1.0
Luteus 2, W	7	8.6	0.26	0.80	43.8	89	96	37	133	22.3	51.2	130	0.9
Luteus 3, W	29	11.8	0.31	0.90	51.6	78	86	41	127	28.9	63.3	161	1.2
Luteus 3, Br	5	12.7	0.38	0.82	49.6	71	80	47	128	33.1	39.6	114	1.1
LSD <i>luteus</i>		1.1	0.07	0.18	4.6	2	2	2	2	5.0	17.2	48	0.2

^aSpecies, Cluster within species, Accession within cluster. Percentage of variance captured by each classification level in nested ANOVA. *** $P < 0.001$.

PC2 (Figure 4; see Table 4 for species and cluster means). As in 2007, balanced within-cluster comparisons confirmed higher vigor/earlier phenology in domesticated compared to wild material in the 2008 data (Table 4, Figure 4: *L. angustifolius*, Cluster 4, and *L. albus*, Cluster 5).

Regression was used to more closely investigate the consistent phenology trade-offs highlighted in PC1 in both 2007 and 2008 ordinations. Early vigor increased in proportion to seed size in a common manner among all 3 species (Figure 5A: $y = 0.07 + 0.02x$, $P < 0.001$). Seed size was negatively related to flowering date, largely decreasing at similar rates between or within species as flowering became later (Figure 5B,

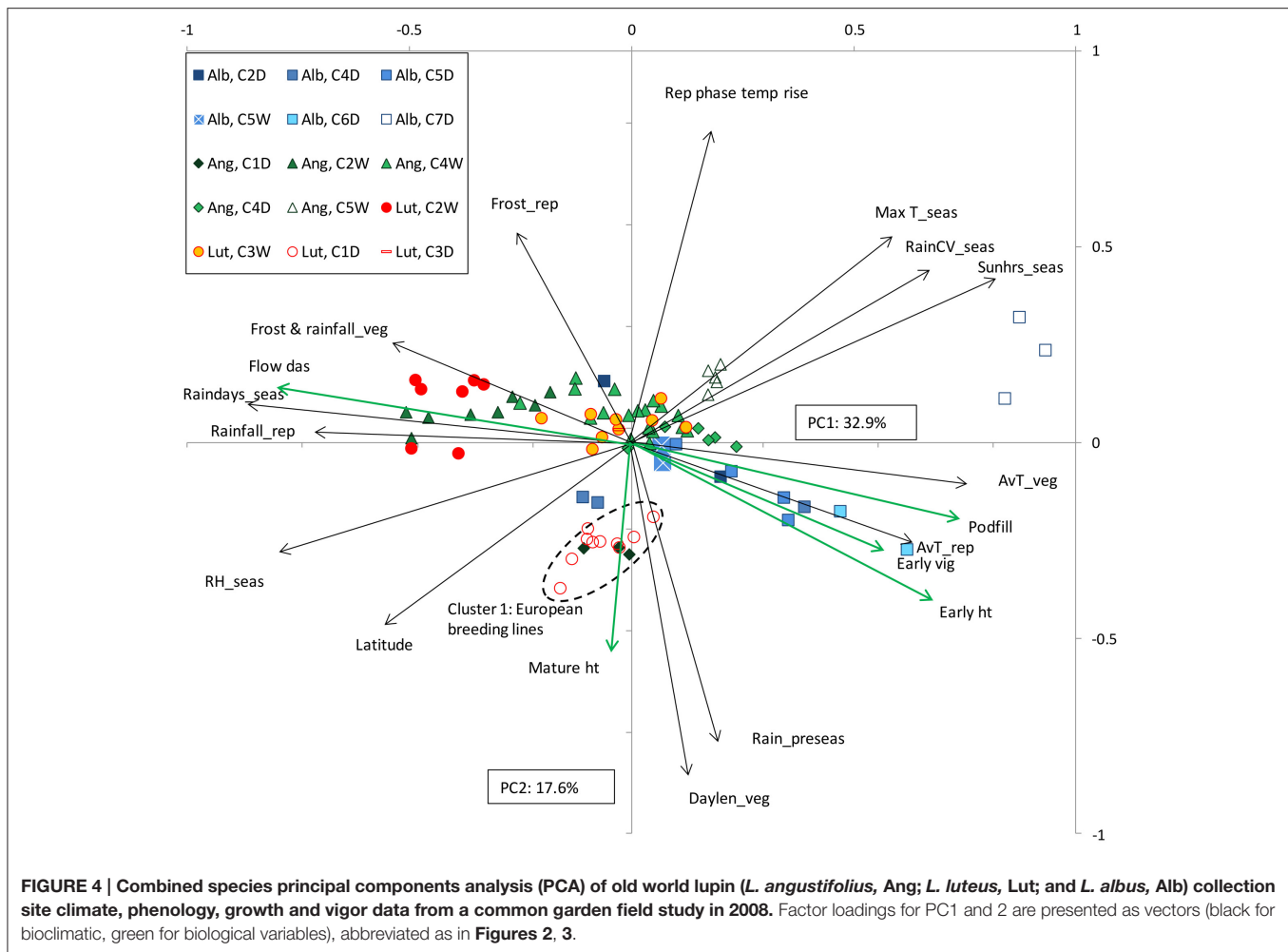
Table 3). Domesticated *L. luteus* and *L. angustifolius* were notable exceptions where there was no significant relationship between seed size and flowering date (Table 3). Wild *L. albus* appeared to be more responsive than the others, while in domesticated, spring-sown *L. albus*, seed size appeared to increase, rather than decrease with flowering time (Figure 5B, Table 3). Note that both these groups were relatively small, and regressions subject to leverage effects. The length of the reproductive phase was also negatively related to flowering date across all species, with the strongest responses in wild and domesticated spring-sown *L. albus* (Figure 5C, Table 3). While winter-sown *L. albus* and *L. luteus* responses were parallel, *L. albus* pod fill



Hard seed breakdown patterns were typically logistic in both *L. luteus* and *L. angustifolius* (**Figure 6**). Unlike the data presented previously, both species showed considerable genotypic variation that was unrelated to rainfall, such that there were no significant differences in both the starting intercepts and rates of hard seed breakdown between low and high rainfall ecotypes. Thus, during the first growing season after seed maturation high rainfall ecotypes could be extremely soft, or completely impermeable to water, and a similar range existed within low rainfall ecotypes (**Figure 6**). This situation appeared to continue into the 2nd year, notwithstanding further loss of physical dormancy in the second summer. Nor were there any relationships between early season soft seed proportion and flowering time or summer temperature mean, maxima or range in either species (*L. luteus*, $r^2 = 0.02\text{--}0.13$, $P = 0.494\text{--}0.959$; *L. angustifolius*, $r^2 = 0.03\text{--}0.08$, $P = 0.123\text{--}0.579$, data not presented). Nevertheless, there were clear cut specific differences: the proportion of soft seed tended to be far higher in *L. luteus* than in *L. angustifolius*. Seed banks of almost all *L. luteus* accessions were >30% soft at harvest, rising >70% during the first growing season. By contrast, the *L. angustifolius* seed bank was typically <10% soft at harvest, rising to <30% during the first growing season, well after the onset of opening rains (**Figure 6**).

DISCUSSION

The role of phenology in avoiding perennial stresses, such as winter cold and terminal drought is well understood in



Mediterranean annuals (Berger et al., 2016), and can lead to the evolution of regionally appropriate control mechanisms (Berger et al., 2011). This lupin example suggests that early phenology may also be important in supporting the production of larger seeds. In a growing season terminated by drought, early flowering allows for long pod filling phases (Figure 5C) that support the production of large seeds, important because these are likely to require more time to fill (Vile et al., 2006). Moreover, a long reproductive phase will reduce fecundity-seed size trade-offs in large seeded genotypes (Norman et al., 2005) by providing more opportunities for seed production. The reverse argument applies equally well to late flowering, high rainfall ecotypes. Here smaller, presumably faster developing seeds (Vile et al., 2006) reduce the fecundity constraints of a shorter reproductive phase.

So why should drought-prone Mediterranean habitats select for larger seeds? Negative correlations between seed size and moisture availability occur across wide-ranging community types in Californian herbs (Baker, 1972), while Mazer (1989) found that early and short duration flowering were associated with the production of large seeds in Indiana dune angiosperms. Similarly, Norman et al. (2005) found negative correlations between seed size and flowering date among 19 *Trifolium* species sourced

from Mediterranean climates in the eastern Mediterranean and southern Australia. The role of seed size in promoting early establishment, vigor and competition in annual species is well known (Leishman et al., 2000). Therefore, under unfavorable conditions (such as aridity or shading) seed size is positively correlated to reproductive yield, trumping the usual fecundity-size trade-off (Venable and Brown, 1988). Rainfall amount and variability are inversely related cross the Mediterranean lupin range: drier sites tend to have more variable, less frequent rainfall (Berger et al., 2008, see also vectors in Figures 2, 4). Accordingly, it is important for low rainfall ecotypes to establish as quickly as possible after the growing season opening rains because the timing of the next rainfall is uncertain. This issue is exacerbated in Old World lupins, which are adapted to poorly fertile, acid sandy soils with little water-holding capacity, where the wetting front is likely to move rapidly downwards (Palta et al., 2012). Under these circumstances, maximizing early vigor in low rainfall ecotypes seems appropriate in the Old World lupins.

Given the potential for the regulation of germination to frame in-season phenology, and expected trade-offs between seed size and dormancy (Venable and Brown, 1988), it was interesting not to find a relationship between hardseed breakdown and

TABLE 4 | Within and between old world lupin species differences in terms of variance distribution^a and mean values for seed size, early growth, phenology and productivity in a rainfed field common garden evaluation in 2008.

Category		Early vigour	Growth rate bio	Growth rate ht	Plant ht	Flower days	Rep phase	Maturity days	Biomass
^a Species variance		93.6***	54.4***	64.2***	58.2***	90.7***	96.7***	95.1***	56.1
^a Cluster within species		4.4***	24.7***	28.5***	29.5***	8.2***	2.4***	3.3***	24.8
^a Accession within cluster		2.0***	20.9***	7.4***	12.3***	1.2***	0.9***	1.6***	19.2
Species means	<i>n</i>								
<i>L. albus</i> L.	17	0.65	0.023	0.041	56	61	103	164	27.9
<i>L. angustifolius</i> L.	41	0.32	0.019	0.035	50	71	81	152	22.5
<i>L. luteus</i> L.	28	0.38	0.018	0.038	50	82	72	153	22.7
LSD Species mean		0.03	0.003	0.004	1.8	0.7	1.4	1.3	2.5
Cluster means									
<i>L. albus</i> L.									
Albus 2, LR	2	0.67	0.026	0.046	62	64	104	167	33.5
Albus 4, LR	3	0.50	0.018	0.040	56	66	103	169	22.7
Albus 5, LR	4	0.71	0.024	0.044	62	60	106	166	28.9
Albus 5, W	2	0.64	0.025	0.046	54	68	91	160	32.2
Albus 6, LR	3	0.57	0.026	0.041	57	61	102	165	35.2
Albus 7, LR	3	0.77	0.016	0.028	43	51	107	158	17.5
LSD <i>albus</i>		0.14	0.010	0.012	6.5	2.1	3.3	3.3	9.9
<i>L. angustifolius</i> L.									
Angus 1, Br	3	0.38	0.022	0.041	58	71	81	152	25.0
Angus 2, W	8	0.24	0.017	0.042	54	77	77	154	20.3
Angus 3, W	2	0.17	0.016	0.033	42	76	75	152	17.6
Angus 4, Br	6	0.38	0.022	0.033	51	66	88	153	25.5
Angus 4, W	18	0.35	0.020	0.034	48	70	81	151	23.8
Angus 5, W	5	0.26	0.014	0.029	42	68	84	153	17.8
LSD <i>angustifolius</i>		0.09	0.008	0.010	6.3	2.4	5.9	5.0	8.7
<i>L. luteus</i> L.									
Luteus 1, Br	10	0.41	0.017	0.036	51	75	76	151	21.8
Luteus 2, W	7	0.27	0.019	0.041	49	95	62	156	22.1
Luteus 3, Br	1	0.36	0.013	0.038	47	79	74	153	13.4
Luteus 3, W	10	0.43	0.020	0.038	49	80	73	153	24.9
LSD <i>luteus</i>		0.12	0.009	0.011	5.9	2.7	4.7	4.2	10.5

^aSpecies, Cluster within species, Accession within cluster. Percentage of variance captured by each classification level in nested ANOVA. ****P* < 0.001.

collection site aridity in the “Old World” lupins (**Figure 6**). Low rainfall collection sites tend to have higher summer temperature maxima and wider diurnal ranges than high rainfall sites (**Table 2**). While there is clearly a strong summer pattern of hardseed breakdown, variation in intercepts and rates of dormancy loss suggest that this process has elements of both predictive and bet-hedging strategies (Donaldson-Matasci et al., 2013). The former protects against untimely germination, while the latter spreads risk over time. Both responses vary widely between *L. luteus* and *L. angustifolius* low and high rainfall ecotypes such that these populations are buffered by a range of germination strategies. Even in strongly Mediterranean climates rainfall variability may be too high to select for a uniform predictive dormancy response, especially given low water holding capacity soils, and the likelihood of other unpredictable disturbance, such as grazing. Moreover, there are

other advantages attributed to diversified bet-hedging related to the effective reduction in seed density associated with staggered germination. These include escaping crowding, reducing sib competition and the potential for inbreeding (Rubio de Casas et al., 2015). Indeed, studies of Mediterranean legumes often return very weak (Piano et al., 1996) or no relationships between hardseed breakdown and flowering time or collection site rainfall (Gladstones, 1967; Norman et al., 2002).

The overlap between natural and human selection in all 3 species is particularly interesting. Domesticated preferences have consistently selected for the “drought-adapted package” even though domestication occurred in very different environments, separated by 1000s of years. *L. albus* was domesticated in the Aegean during the Bronze Age (Zohary and Hopf, 2000); while *L. luteus* and *L. angustifolius* were domesticated as spring-sown temperate European crops in the 18–19th centuries

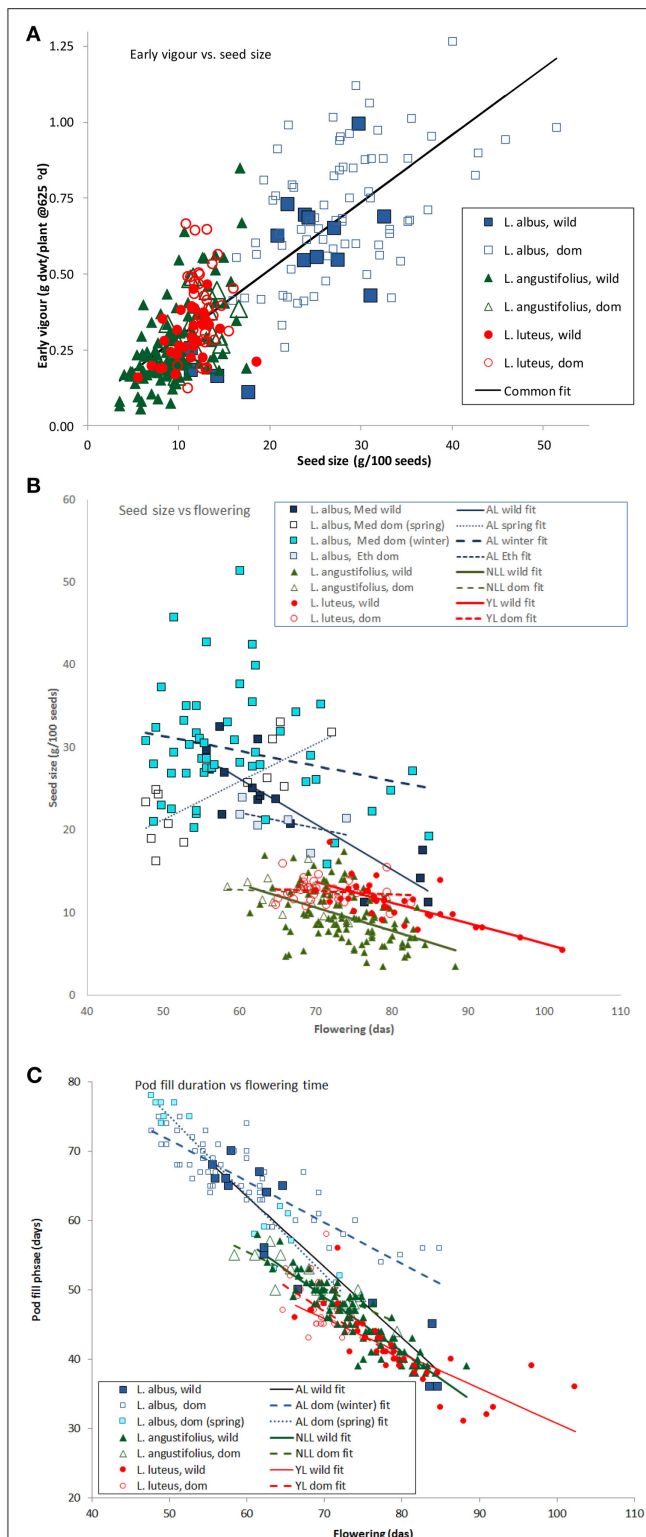


FIGURE 5 | The effects of seed size on early vigor (A), and flowering date on seed size (B), and the length of the reproductive phase (C) in old world lupins (data from a 2007 common garden field trial). In (A) a common linear regression accounted for 66.2% of variance, while in (B,C) separate regression equations for species accounted for 82.5 and 93.1%, respectively. (Regression equations for (B,C) presented in Table 5).

(Hondelmann, 1984), moving to Mediterranean Australia in the late twentieth century (Gladstones, 1994). *L. albus* is traditionally grown for human consumption as a whole seed where larger seeds are preferred (Huyghe, 1997), therefore seed size likely to have been under direct selection. Moreover, as a Mediterranean crop planted after the season opening rains (or indeed considerably later, as in the case of Balkan spring-sown crops; Mihailovic et al., 2008) its development is delayed compared to those hard-seeded wild populations germinating on soil moisture, exerting strong selection pressure for early vigor/phenology to catch up and complete the lifecycle prior to the onset of terminal drought. The European domestications selected for similar trait combinations, but for different reasons. Early attempts to domesticate *L. albus* failed due to the inability of the crop to mature in the absence of terminal drought (Hondelmann, 1984). Subsequent efforts with *L. luteus* and *L. angustifolius* selected strongly on phenology and early vigor, attested to by cultivar names such as Pflugs Allerfrühste (plow's earliest) (Gladstones, 1970), and reflected in the early, flowering-unresponsive maturity of domesticated material in this study. Later Australian efforts in warm, short season Mediterranean environments greatly enhanced this prior selection for earliness, selecting for extremely temperature responsive phenology (Berger et al., 2012a,b), equivalent to that of Southern Indian chickpea, an environment $<10^{\circ}\text{C}$ warmer than the northern Western Australian grainbelt (25.8° vs. 14.5°C) (Berger et al., 2011).

Having discussed similarities among the 3 Old World lupin species, it is important to consider the differences. While all species delay flowering with increasing collection site rainfall, *L. luteus* is particularly responsive, and comes from a late flowering baseline (intercept). Even low rainfall ecotypes (Cluster 3) are relatively late, a phenology that is delayed even more in high rainfall environments such as northern Iberia (Figure 2B). This is complemented by a greater tendency for soft seededness, such that a large proportion will germinate very early, perhaps allowing late flowering *L. luteus* to transition to reproduction at the same time as the much harder seeded *L. angustifolius*. In high rainfall ecotypes of *L. luteus*, the combination of early germination and late phenology drives a very long vegetative phase. This underpins massive above and below-ground biomass development, driving very high rates of water-use compared to earlier flowering, lower biomass low rainfall ecotypes (Berger and Ludwig, 2014). We assume that this aggressive strategy increases the competitive capacity of high rainfall *L. luteus* ecotypes, as high root biomass facilitates exploitation of nutrients and water, while large leaf area provides the C fixation to drive prolonged high growth rates, both of which are complemented by early establishment (Gremer et al., 2016). However, high water use leads to early stress onset under terminal drought. This is partly mitigated in high rainfall ecotypes only, by the generation of lower critical leaf water potentials and maintenance of higher relative water contents (Berger and Ludwig, 2014). Clearly there are limits to this drought tolerance capacity, as evidenced by low productivity of high rainfall ecotypes in the terminal drought stressed 2007 common garden. Accordingly, the early emergence, long vegetative phase, high biomass, resource acquisition and

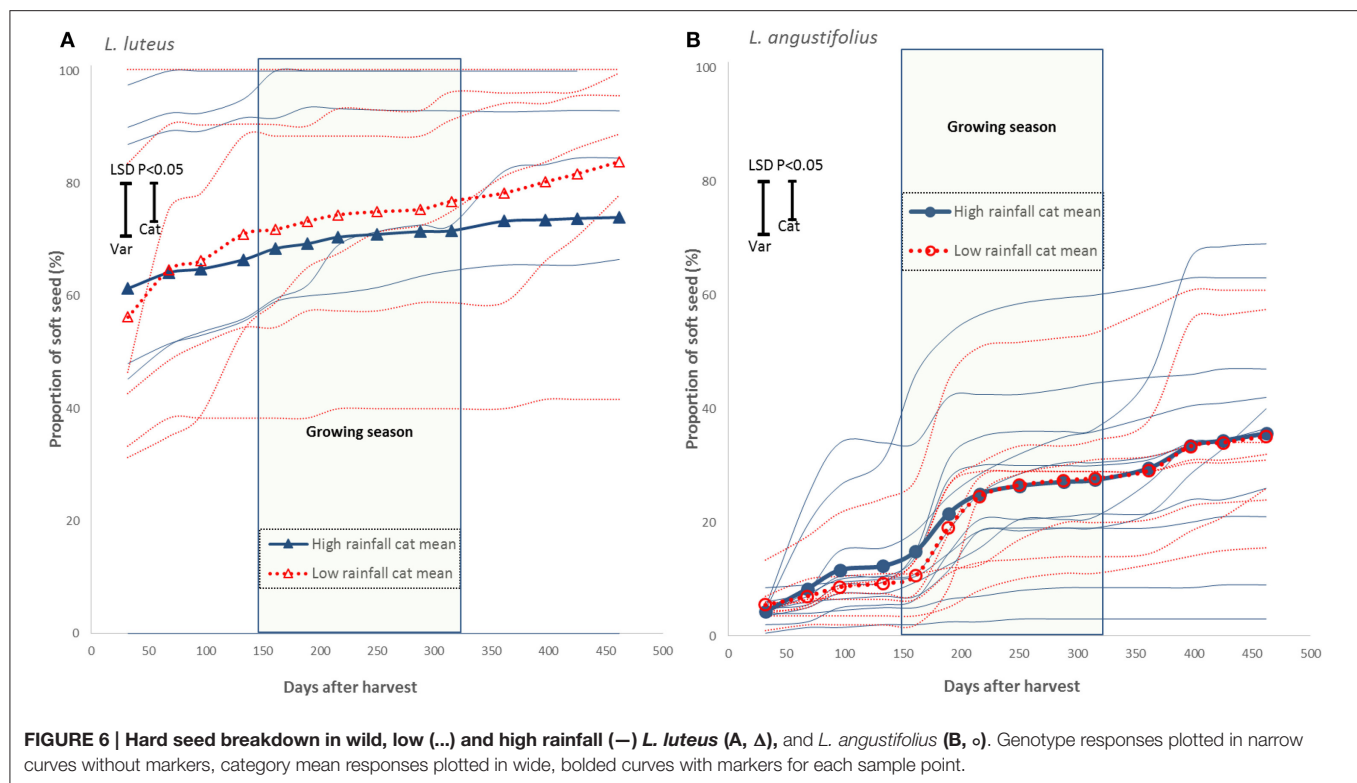


TABLE 5 | Linear equations for relationships between flowering, maturity, seed size and pod fill in Old World lupin groups regressed in Figures 2B,C, 5B,C.

Species	Flowering-rainfall (Figure 2B)		Maturity-flowering (Figure 2C)		Flower-seed size (Figure 5B)		Pod fill-flower (Figure 5C)	
	Intercept	Slope	Intercept	Slope	Intercept	Slope	Intercept	Slope
<i>L. albus</i> , dom (winter)	51.4***	0.015***	113.8***	0.41***	40.4***	-0.18**	101.4***	-0.60***
<i>L. albus</i> , dom (spring)	51.4***	0.015***	145.9***	-0.15 (=0.09)	-1.5NS	0.46***	132.3***	-1.15***
<i>L. albus</i> , wild	51.4***	0.015***	146.1***	-0.17*	59.1***	-0.55***	124.7***	-1.02***
<i>L. angustifolius</i> , dom	68.0***	0.009***	108.9***	0.27*	15.4NS	-0.04NS	87.2***	-0.53***
<i>L. angustifolius</i> , wild	68.0***	0.009***	117.3***	0.13**	30.3***	-0.28***	104.3***	-0.79***
<i>L. luteus</i> , dom	66.7***	0.18***	120.6***	0.08NS	15.4NS	-0.04NS	96.8***	-0.71***
<i>L. luteus</i> , wild	66.7***	0.18***	96.4***	0.40***	31.3***	-0.25**	81.1***	-0.50***
LSD ($P < 0.05$)	3.4	0.005	10.4	0.15	17.5	0.26	11.0	0.16

P-values present outcomes from *T*-tests of H_0 intercept and slope = 0: NS > 0.05, * < 0.05, ** < 0.01, *** < 0.001.

water-use lifecycle of *L. luteus* is a high risk, competitive strategy for specific adaptation to longer season, high rainfall environments, which may explain the scarcity of the species in the more drought-prone southern Mediterranean (Figure 1). Certainly, *L. luteus* has the narrowest, most coastal distribution of the 3 Old world species in the present study (Gladstones, 1998). Yellow lupin has struggled to establish itself as a break crop in Mediterranean agro-ecosystems, and our results suggest that re-imagining its role in the system is warranted. Perhaps there is a case for an extremely early-sown, drought tolerant, late variety for Mediterranean high rainfall zones?

Conversely, *L. angustifolius* is widely distributed around the Mediterranean (Gladstones, 1998). High rainfall ecotypes appear

to be much more conservative than in *L. luteus*, with a higher proportion of hard seed, flowering and maturing much earlier, with a relatively rainfall-unresponsive phenology (Figures 2B,C). These characteristics identify *L. angustifolius* as a typical drought escaping species (Berger et al., 2016), and therefore we wonder to what extent there are adaptive traits that can be exploited for higher rainfall environments, such as the competitive, but risky, high productivity strategy of *L. luteus*. Currently, there are no elite narrow leafed lupin cultivars targeting long season, high yield potential environments (Berger et al., 2012a; Chen et al., 2016).

Lupinus albus is very different: flowering earlier but maturing much later than *L. luteus* and *L. angustifolius*; leading to longer

reproductive phases that support much larger seed sizes and early vigor. This ruderal/competitive combination appears to give *L. albus* a very broad adaptive capacity, reflected in its relatively wider Mediterranean and North African distribution (Figure 1; Gladstones, 1998). How *L. albus* combines early flowering and late maturity is unknown. Perhaps its greater early vigor allows the species to access more water from lower in the soil profile than *L. luteus* and *L. angustifolius*, sustaining a longer reproductive phase under terminal drought. *L. albus* seedlings certainly accumulate much greater lateral root numbers, total root length and biomass, at higher length to weight ratios than either *L. luteus* or *L. angustifolius* (Clements et al., 1993). Alternatively, the species may be less profligate than the others in its vegetative water-use, saving water for the reproductive phase. This strategy has been documented in crops adapted to fine-textured, higher water-holding capacity soils, such as chickpea and pearl millet (Kholová et al., 2010; Zaman-Allah et al., 2011), but is harder to credit in such a vigorous, coarse-textured soil adapted species.

We conclude that while trade-offs between phenology, seed size and vigor operate in a similar manner in these 3 Old World lupin species as they transition from low to high rainfall environments, or from wild Mediterranean annual to

domesticated crop, the combination of these and associated traits defines their adaptive potential, and is reflected in their natural distribution. Given that *L. luteus*, *L. angustifolius* and even *L. albus* are still relatively unimproved, partly domesticated crops, plant breeders can exploit these relationships to increase yield in target environments.

AUTHOR CONTRIBUTIONS

JB designed helped implement the experiments, analyzed the results, and wrote the paper. DS, CL implemented the experiments, and provided feedback on the paper.

ACKNOWLEDGMENTS

The authors would like to acknowledge generous research funding support from the Grains Research and Development Corporation, Australia and the Commonwealth Scientific and Industrial Research Organisation (CSIRO). The Department of Agriculture and Food (Western Australia, DAFWA) is thanked for providing both the passport data and genetic resources for this study.

REFERENCES

- Abbo, S., Berger, J., and Turner, N. C. (2003). Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.* 30, 1081–1087. doi: 10.1071/FP03084
- Baker, H. G. (1972). Seed Weight in relation to environmental conditions in California. *Ecology* 53, 997–1010. doi: 10.2307/1935413
- Berger, J. D., Buirchell, B., Luckett, D. J., and Nelson, M. N. (2012a). Domestication bottlenecks limit genetic diversity and constrain adaptation in narrow-leaved lupin (*Lupinus angustifolius* L.). *Theor. Appl. Genet.* 124, 637–652. doi: 10.1007/s00122-011-1736-z
- Berger, J. D., Buirchell, B., Luckett, D. J., Palta, J. A., Ludwig, C., and Liu, D. (2012b). How has narrow-leaved lupin changed in its 1st 40 years as an industrial, broad-acre crop? A GxE-based characterization of yield-related traits in Australian cultivars. *Field Crops Res.* 126, 152–164. doi: 10.1016/j.fcr.2011.10.006
- Berger, J. D., Clements, J. C., Nelson, M. N., Kamphuis Lars, G., Singh Karam, B., and Buirchell, B. (2013). The essential role of genetic resources in narrow-leaved lupin improvement. *Crop Pasture Sci.* 64, 361–373. doi: 10.1071/CP13092
- Berger, J. D., and Ludwig, C. (2014). Contrasting adaptive strategies to terminal drought stress gradients in Mediterranean legumes: phenology, productivity and water relations in wild and domesticated *Lupinus luteus* L. *J. Exp. Bot.* 65, 6219–6229. doi: 10.1093/jxb/eru006
- Berger, J. D., Ludwig, C., and Buirchell, B. J. (2008). “Ecogeography of the old world lupins: characterising the habitat range,” in *Lupins for Health and Wealth. Proceedings of the 12th International Lupin Conference, Fremantle, Western Australia*, 14–18 September 2008, eds J. A. Palta and J. D. Berger (Canterbury: International Lupin Association), 355–361.
- Berger, J. D., Milroy, S. P., Turner, N. C., Siddique, K. H. M., Imtiaz, M., and Malhotra, R. (2011). Chickpea evolution has selected for contrasting phenological mechanisms among different habitats. *Euphytica* 180, 1–15. doi: 10.1007/s10681-011-0391-4
- Berger, J., Palta, J., and Vadez, V. (2016). Review: an integrated framework for crop adaptation to dry environments: responses to transient and terminal drought. *Plant Sci.* 253, 58–67. doi: 10.1016/j.plantsci.2016.09.007
- Chen, C., Berger, J., Fletcher, A., Lawes, R., and Robertson, M. (2016). Genotype × environment interactions for phenological adaptation in narrow-leaved lupin: a simulation study with a parameter optimized model. *Field Crops Res.* 197, 28–38. doi: 10.1016/j.fcr.2016.08.002
- Clements, J. C., White, P. F., and Buirchell, B. J. (1993). The root morphology of *Lupinus angustifolius* in relation to other *Lupinus* species. *Aust. J. Agric. Res.* 44, 1367–1375. doi: 10.1071/AR931367
- Cocks, P. S. (1999). “Reproductive strategies and genetic structure of wild and naturalised legume populations,” in *Genetic Resources of Mediterranean Pasture and Forage Legumes*, Vol. 33, eds S. J. Bennett and P. S. Cocks (Netherlands: Springer), 20–31.
- Donaldson-Matasci, M. C., Bergstrom, C. T., and Lachmann, M. (2013). When unreliable cues are good enough. *Am. Nat.* 182, 313–327. doi: 10.1086/671161
- Gladstones, J. (1967). Naturalized subterranean clover strains in Western Australia: a preliminary agronomic examination. *Aust. J. Agric. Res.* 18, 713–731.
- Gladstones, J. S. (1970). Lupins as crop plants. *Field Crop Abstr.* 23, 123–148.
- Gladstones, J. S. (1974). *Lupins of the Mediterranean Region and Africa*. Perth, WA: Department of Agriculture Technical Bulletin No. 26.
- Gladstones, J. S. (1994). “An historical review of lupins in Australia,” in *1st Lupin Technical Symposium*, eds M. Dracup and J. A. Palta (Perth, WA: Department of Agriculture), 1–38.
- Gladstones, J. S. (1998). “Distribution, origin, taxonomy, history and importance,” in *Lupins as crop plants: Biology, Production and Utilization*, eds J. S. Gladstones, C. A. Atkins and J. Hamblin (Wallingford: CAB International), 1–37.
- Gremer, J. R., Kimball, S., and Venable, D. L. (2016). Within- and among-year germination in Sonoran Desert winter annuals: bet hedging and predictive germination in a variable environment. *Ecol. Lett.* 19, 1209–1218. doi: 10.1111/ele.12655
- Grime, J. P. (1979). *Plant Strategies and Vegetation Processes*. (Chichester: Wiley & Sons).
- Hacker, J. B. (1984). Genetic variation in seed dormancy in digitaria milaniana in relation to rainfall at the collection site. *J. Appl. Ecol.* 21, 947–959. doi: 10.2307/2405059
- Hacker, J. B., and Ratcliff, D. (1989). Seed dormancy and factors controlling dormancy breakdown in buffel grass accession from contrasting provenances. *J. Appl. Ecol.* 26, 201–212.

- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* 25, 1965–1978. doi: 10.1002/joc.1276
- Hondelmann, W. (1984). The Lupin-ancient and modern crop plant. *Theor. Appl. Genet.* 68, 1–9.
- Huyghe, C. (1997). White lupin (*Lupinus albus* L.). *Field Crops Res.* 53, 147–160.
- Kato, K., Tanizoe, C., Beiles, A., and Nevo, E. (1998). Geographical variation in heading traits in wild emmer wheat, *Triticum dicoccoides*. II. Variation in heading date and adaptation to diverse eco-geographical conditions. *Hereditas* 128, 33–39.
- Kholová, J., Hash, C. T., Kumar, P. L., Yadav, R. S., Kočová, M., and Vadez, V. (2010). Terminal drought-tolerant pearl millet [*Pennisetum glaucum* (L.) R. Br.] have high leaf ABA and limit transpiration at high vapour pressure deficit. *J. Exp. Bot.* 61, 1431–1440. doi: 10.1093/jxb/erq013
- Leishman, M. R., Wright, I. J., Moles, A. T., and Westoby, M. (2000). “The evolutionary ecology of seed size,” in *Seeds: The Ecology of Regeneration in Plant Communities*, ed M. Fenner (Wallingford: CABI Publisher), 31–57.
- Mazer, S. J. (1989). Ecological, taxonomic, and life history correlates of seed mass among indian dune angiosperms. *Ecol. Monogr.* 59, 153–175. doi: 10.2307/2937284
- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Mihailovic, V., Hill, G. D., Mikic, A., Cupina, B., and Vasiljevic, S. (2008). “White lupin as a forage crop on alkaline soils,” in *Lupins for Health and Wealth. Proceedings of the 12th International Lupin Conference, Fremantle, Western Australia, 14-18 September 2008*, eds J. A. Palta and J. D. Berger (Canterbury: International Lupin Association), 79–82.
- Mülayim, M., Tamkoç, A., and Babaoglu, M. (2002). Sweet white lupins versus local bitter genotype: agronomic characteristics as affected by different planting densities in the Gölle region of Turkey. *Eur. J. Agron.* 17, 181–189. doi: 10.1016/S1161-0301(02)00007-2
- Norman, H. C., Cocks, P. S., and Galwey, N. W. (2002). Hardseededness in annual clovers: variation between populations from wet and dry environments. *Aust. J. Agric. Res.* 53, 821–829. doi: 10.1071/AR01115
- Norman, H. C., Cocks, P. S., and Galwey, N. W. (2005). Annual clovers (*Trifolium* spp.) have different reproductive strategies to achieve persistence in Mediterranean-type climates. *Aust. J. Agric. Res.* 56, 33–43. doi: 10.1071/AR03236
- Palta, J. A., Berger, J. D., and Bramley, H. (2012). “Physiology of the yield under drought: lessons from studies with lupin,” in *Plant Responses to Drought Stress: From Morphological to Molecular Features*, ed R. Aroca (New York, NY: Springer Verlag), 417–440.
- Pannell, D. J. (1995). Economic aspects of legume management and legume research in dryland farming systems of southern Australia. *Agric. Syst.* 49, 217–236. doi: 10.1016/0308-521X(94)00049-W
- Piano, P., Pecetti, L., and Carroni, A. M. (1996). Climatic adaptation in subterranean clover populations. *Euphytica* 92, 39–44. doi: 10.1007/BF00022826
- Rubio de Casas, R., Donohue, K., Venable, D. L., and Cheptou, P.-O. (2015). Gene-flow through space and time: dispersal, dormancy and adaptation to changing environments. *Evol. Ecol.* 29, 813–831. doi: 10.1007/s10682-015-9791-6
- Thompson, J. D. (2005). “Trait variation, adaptation, and dispersal in the Mediterranean mosaic,” in *Plant Evolution in the Mediterranean* (Oxford: OUP Oxford), 109–167.
- Venable, D. L., and Brown, J. S. (1988). The selective interactions of dispersal, dormancy, and seed size as adaptations for reducing risk in variable environments. *Am. Nat.* 131, 360–384. doi: 10.1086/284795
- Vile, D., Shipley, B., and Garnier, E. (2006). A structural equation model to integrate changes in functional strategies during old-field succession. *Ecology* 87, 504–517. doi: 10.1890/05-0822
- Volis, S. (2007). Correlated patterns of variation in phenology and seed production in populations of two annual grasses along an aridity gradient. *Evol. Ecol.* 21, 381–393. doi: 10.1007/s10682-006-9108-x
- Yousfi, N., Slama, I., Ghnaya, T., Sayoure, A., and Abdely, C. (2010). Effects of water deficit stress on growth, water relations and osmolyte accumulation in *Medicago truncatula* and *M. lacinata* populations. *Comp. Rendus Biol.* 333, 205–213. doi: 10.1016/j.crv.2009.12.010
- Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011). Chickpea genotypes contrasting for seed yield under terminal drought stress in the field differ for traits related to the control of water use. *Funct. Plant Biol.* 38, 270–281. doi: 10.1071/FP10244
- Zohary, D., and Hopf, M. (2000). “Lupinus,” in *Domestication of Plants in the Old World* (Oxford: Clarendon Press), 122–124.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Berger, Shrestha and Ludwig. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genotypic Variation in a Breeding Population of Yellow Sweet Clover (*Melilotus officinalis*)

Kai Luo^{1†}, M. Z. Z. Jahufer^{2†}, Fan Wu¹, Hongyan Di¹, Daiyu Zhang¹, Xuanchen Meng¹, Jiyu Zhang^{1*} and Yanrong Wang^{1*}

¹ State Key Laboratory of Grassland Agro-Ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, ² AgResearch Limited, Grasslands Research Centre, Palmerston North, New Zealand

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Institute for Sustainable
Agriculture - CSIC, Spain

Reviewed by:

Inger Martinussen,
Norwegian Institute of Bioeconomy
Research, Norway
Juan Marcelo Zabala,
Universidad Nacional del Litoral,
Argentina

*Correspondence:

Jiyu Zhang
zhangjiy@lzu.edu.cn;
Yanrong Wang
yrwang@lzu.edu.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 16 December 2015

Accepted: 20 June 2016

Published: 12 July 2016

Citation:

Luo K, Jahufer MZZ, Wu F, Di H,
Zhang D, Meng X, Zhang J and
Wang Y (2016) Genotypic Variation in
a Breeding Population of Yellow Sweet
Clover (*Melilotus officinalis*).
Front. Plant Sci. 7:972.
doi: 10.3389/fpls.2016.00972

Yellow sweet clover is a widely spread legume species that has potential to be used as a forage crop in Western China. However, limited information is available on the genetic variation for herbage yield, key morphological traits, and coumarin content. In this study, 40 half sib (HS) families of *M. officinalis* were evaluated for genotypic variation and phenotypic and genotypic correlation for the traits: LS (leaf to stem ratio), SV (spring vigor), LA (leaf area), PH (plant height), DW (herbage dry weight), SD (stem diameter), SN (stem number), Cou (coumarin content), SY (seed yield), across two locations, Yuzhong and Linze, in Western China. There was significant ($P < 0.05$) genotypic variation among the HS families for all traits. There was also significant ($P < 0.05$) genotype-by-environment interaction for the traits DW, PH, SD, SN, and SV. The estimates of HS family mean repeatability across two locations ranged from 0.32 for SN to 0.89 for LA. Pattern analysis generated four HS family groups where group 3 consisted of families with above average expression for DW and below average expression for Cou. The breeding population developed by polycrossing the selected HS families within group 3 will provide a significant breeding pool for *M. officinalis* cultivar development in China.

Keywords: forage breeding, genotypic variation, genotype-by-environment interactions, correlation coefficient, coumarin

INTRODUCTION

Yellow sweet clover known as field melilot or yellow melilot, is an annual or biennial herb that belongs to the Fabaceae family. It is native to temperate and tropical Asia, and Europe (GRIN, 2000). *Melilotus officinalis* is one of the most common species in the *Melilotus* genus. This species has adaptation to environmental constraints such as drought and cold (Turkington et al., 1978) and salinity (Sherif, 2009). *Melilotus* is used as a ground cover in depleted soils (Allen and Allen, 1981), especially in moderately saline areas where traditional forage legumes cannot be grown (Maddaloni, 1986). *Melilotus officinalis* usually occurs in the northern region of China, where it is used as green manure for soil fertility improvement and also as a medicinal plant.

Species of *Melilotus*, including yellow sweet clover, have not been widely used in forage production due to their high coumarin content. Coumarin, a secondary plant metabolite, is associated with dicoumarol production. Dicoumarol is an anticoagulant that can cause a haemorrhagic condition known as sweet clover disease (Evans and Kearney, 2003;

Nair et al., 2010). Therefore, the success of forage cultivar development based on any of the *Melilotus* species will depend on a combination of increasing dry matter production and decreasing coumarin content. A number of cultivars of *Melilotus* have been released to date; Acuma, Cumino, Denta, Polara (Smith and Gorz, 1965; Goplen, 1971) for *M. albus* and Norgold (Goplen, 1981), N28, N29 (Gorz et al., 1992) for *M. officinalis*. The *Melilotus* breeding program at Lanzhou University is specifically focused on the development of new cultivars with adaptation to the vast temperate grazing environments of China (Luo et al., 2014).

In any plant breeding program, the rate of genetic gain depends on the genetic diversity for a given trait in the breeding population (Hallauer and Miranda, 1981). Information on the magnitude of genetic variation for key plant attributes in breeding programs will enhance the development of appropriate breeding strategies to achieve maximum genetic gain (Moll and Stuber, 1974). Jahufer and Casler (2015) evaluated the relative merit in genetic gain using single trait selection, correlated response to selection and index selection, based on estimated genetic variation for a range of morphological and quality traits in switch grass (*Panicum virgatum* L.). Genetic variation for key traits have been reported for some of the important forage grasses and legumes: ryegrass (Breese and Hayward, 1972), tall fescue (Piano et al., 2007), white clover (Jahufer et al., 2002), alfalfa (Riday and Brummer, 2007).

There is a lack of quantitative genetic information for *Melilotus*. Few studies have been carried out on the genetic variation for agronomic traits in *Melilotus* species (Ivanov and Chetvertnykh, 1980; Sagalbekov, 1980). Nair et al. (2010) reported genotypic variation for coumarin content among 149 accessions of 15 *Melilotus* species. This study demonstrated the presence of potential genetic variation for coumarin content in *Melilotus* germplasm useful for breeding. However, breeding *Melilotus* species as a forage crop needs to focus on not only coumarin content but also biomass and associated traits. There is also a lack of information on the magnitude of genotype-by-environment interaction effects in *Melilotus*, which will be important for breeding for broad adaptation (Cooper et al., 1993b).

The objective of our study was to conduct a preliminary assessment of the performance of half sib (HS) families of *Melilotus officinalis* across two contrasting locations to: (a) estimate genotypic variation for key traits, and (b) identify families with a combination of superior agronomic performance and low coumarin expression in comparison to two commercial controls.

MATERIALS AND METHODS

Plant Material

Six germplasm accessions (PI 552553 and PI 552554, PI 595394, PI 634019, Ames 22891, and Ames 25658) were selected from a set of 51 accessions that were evaluated for biomass production, agronomy and low coumarin, in Yuzhong, Gansu Province, during 2012–2013 (results not presented). Elite genotypes representing each of the germplasm accessions were polycrossed

in isolation, using honey bees, to generate a breeding population to be used for cultivar development. A total of 40 HS families were generated by harvesting each of the genotypes individually. All harvested seeds from the individual genotypes were kept separately as individual HS families.

Field Trials

The *M. officinalis* HS families were established at two locations: Yuzhong (104°09' E, 35°89' N, elevation 1 653 m a.s.l.) and Linze (100°02' E, 39°15' N, elevation 1 390 m a.s.l.) in Gansu Province, China. There are different climate conditions between Yuzhong and Linze. Yuzhong in Loess Plateau region is a medium temperate semi-arid climate, whereas Linze in the Hexi Corridor is typical desert climate and characterized by an arid climate (Su et al., 2007; Hu et al., 2012; Li et al., 2014). The average annual precipitation in Yuzhong is 295 mm and in Linze is only 117 mm. The mean monthly minimum and maximum temperatures, and total monthly rainfall during the trial period at the two locations are shown in Figure 1.

The soil type at each location is loessal soil at Yuzhong and meadow soil at Linze. The saline-alkali degree was much higher in Linze than in Yuzhong, the salinity is 1.8 ppt in Linze and 0.5 ppt in Yuzhong. Initial soil conditions in Yuzhong and Linze are: pH 7.0 and 7.5, total N of 0.756 g/kg and 0.803 g/kg, total P of 0.752 and 0.708 g/kg, respectively.

At each location, the experimental plots were arranged in a randomized complete block design containing three replicates. Each replicate consisted of the 40 HS families the six parental germplasm accessions and two commercial checks. The origins of these entries are provided in Table 1. The two trials were sown in 15–18 June 2014. The experimental plot size for each entry was 2.4 m² (0.8 × 3 m). Within each plot, the seed was planted at a spacing of 30 cm within-rows and 60 cm between-rows. The plots were fertilized with 150 kg (NH₄)₂HPO₄ ha⁻¹ after sowing.

Measurements

The traits measured were: LS, leaf to stem ratio; SV, spring vigor; LA, leaf area (cm²); PH, plant height (cm); DW, herbage dry weight (g/plant); SD, stem diameter (cm); SN, stem number; Cou, coumarin (% of dry matter); SY, seed yield (g/plant). All the traits were measured in the second year (2015).

Visual scoring for SV was based on a scale of 1 to 5 (1 = low; 5 = high). The morphological traits (PH, SN, SD, and LA) were measured at the flowering stage (50% of the plants had open flowers), resulting in a minimum of three individuals per replicate. LA was measured from three middle leaflets per plant by using a flatbed scanner (EPSON GT-15000) and a WinSEEDLE 2011 image analysis system (Regent Instruments Inc.). Individual plant was harvested for DW measurement at the flowering stage after measuring morphological traits. At harvest, three randomly sampled plants from each replicate were cut off at 3 cm above the soil, placed in paper bags and dried at room temperature (about 20–25°C) with good ventilation until no change in weight was recorded. After measuring DW, the dried samples were hand separated into leaf blade and stem (including the inflorescence and leaf sheath) components and weighted to determine the LS ratio. Three sub samples from each

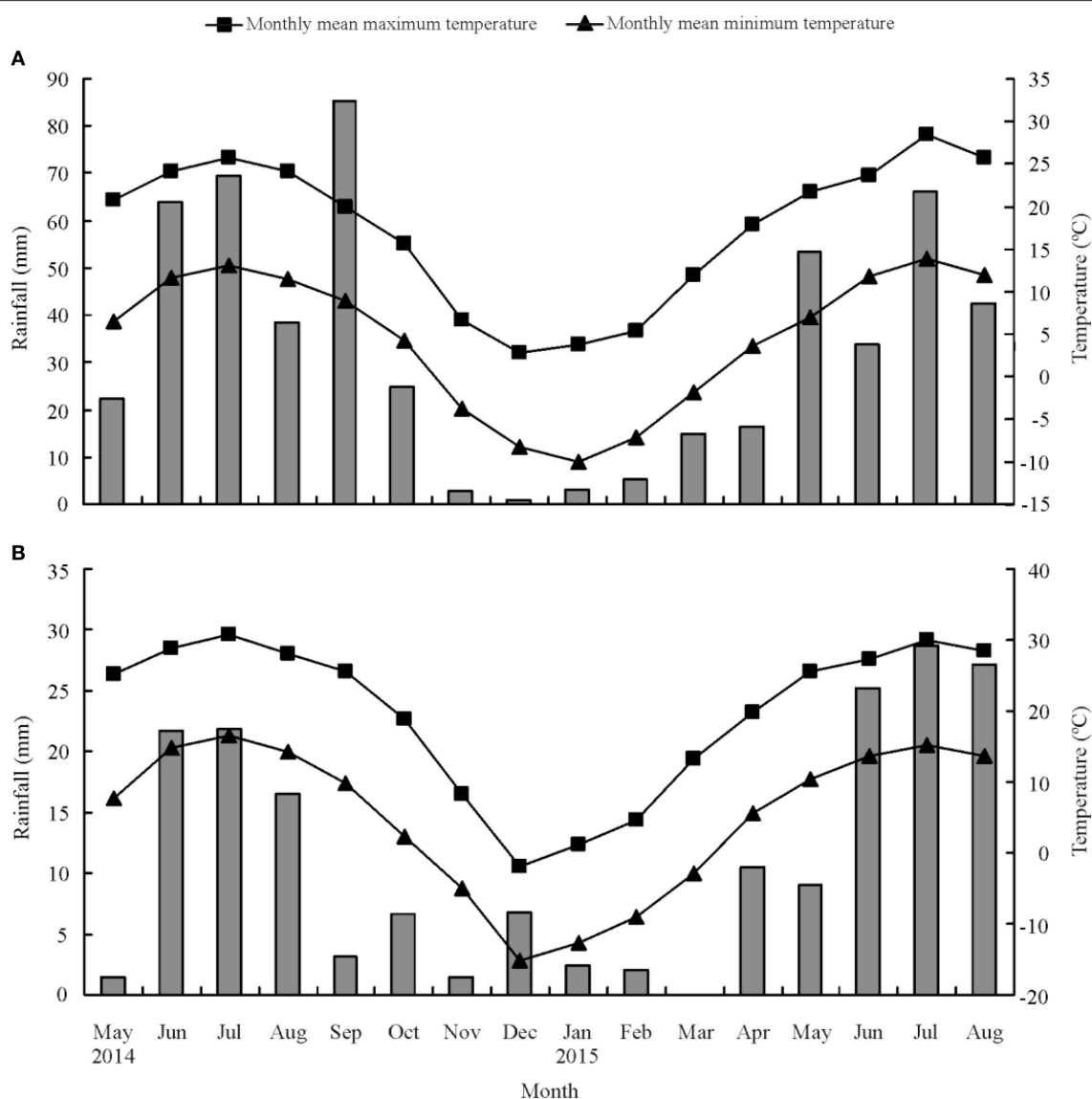


FIGURE 1 | Mean monthly maximum and minimum temperatures (°C) and total monthly rainfall (mm) at Yuzhong (A) and Linze (B), respectively.

field replicate at Yuzhong were combined and ground in a mill to pass through a 1 mm screen for Cou determination. SY was determined from two randomly sampled individuals taken from each replicate when 90% of the pods turned blackish brown at the Linze field. Cou was quantified using HPLC (high performance liquid chromatography, Agilent 1100 series) with a mobile phase of methanol-water (65:35) through an Agilent-XDB C18 column (Zhu and Fan, 2008).

Analysis of Variance

The data were analyzed within and across the two locations Yuzhong and Linze. The analysis across locations was conducted: (a) on only the 40 HS families to estimate genotypic variation, and (b) using all entries in the trial that consisted of the 40 HS families, the six parental germplasm accessions and two

check cultivars, which enabled comparison of progeny, parents, and the commercial material. The analysis was conducted using the variance component analysis procedure, Residual Maximum Likelihood (REML) option, in GenStat 7.1 (2003). A mixed linear model was used for the analyses across the two locations using the REML algorithm.

The linear model used in the analysis was,

$$Y_{ijk} = M + g_i + l_j + r_{jk} + (gl)_{ij} + \varepsilon_{ijk}, \quad (1)$$

Where, Y_{ijk} is the value of an attribute measured from HS family i in replicate k in location j , and $I = 1, \dots, n_g$, $j = 1, \dots, n_l$, $k = 1, \dots, n_r$; M is the overall mean; g_i is the random genotypic effect of HS family i , $N(0, \sigma^2_g)$; l_j is the fixed effect of location j , $N(0, \sigma^2_l)$; r_{jk} is the random effect of replicate k within location j , $N(0, \sigma^2_b)$; $(gl)_{ij}$

TABLE 1 | Origin of the *M. officinalis* germplasm accessions and commercial check cultivars.

Accession number	Source of seed	Latitude and Longitude
PI 552553	Nebraska, United States	41°29'N, 99°54'W
PI 552554	Nebraska, United States	Unknown
Ames 22891	Russia	44°28'N, 40°48'E
Ames 25658	Tien Shan Mountains, China	43°46'N, 89°27'E
PI 595394	Canada	56°07'N, 106°20'W
PI 634019	Saskatchewan, Canada	52°56'N, 106°27'W
Norgold	Nebraska, United States	Commercial cultivar
LX 03	Qinghai, China	Experimental cultivar

is the effect between HS family i and environment j , $N(0, \sigma_{gl}^2)$; ε_{ijk} is the residual effect for HS family i in replicate k in location j , $N(0, \sigma_{\varepsilon}^2)$.

The mixed model analysis generated HS family means based on Best Linear Unbiased Predictors (BLUP) (White and Hodge, 1989). These BLUP values were used to construct a HS family \times trait mean matrix adjusted of HS family \times location interaction effects.

Genotypic Variation and Repeatability

Variation among HS families generated from a population that has gone through at least two cycles of random mating, is an estimate of $\frac{1}{4}$ additive variation of the random mating population they represent (Falconer, 1989). In our study, the 40 HS families were a result of the first random mating of selected germplasm and therefore represented only the F1 generation. Therefore, we do not refer to the variation estimated among the 40 HS families as $\frac{1}{4}$ additive variation, but as genotypic variation, due to a possible combination of additive and non-additive effects. The genotypic variation for the different traits enabled calculation of repeatability, an estimation of the upper limits of their degrees of genetic determination (Falconer, 1989).

The genotypic variance components generated from the REML analysis within and across locations were used to calculate repeatability (R) (Fehr, 1987).

HS family mean repeatability at a single site:

$$R_1 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{\varepsilon}^2}{n_r}} \quad (2)$$

HS family mean repeatability across locations:

$$R_2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_{gl}^2}{n_l} + \frac{\sigma_{\varepsilon}^2}{n_l n_r}} \quad (3)$$

Where, in both model (2) and model (3), the respective variance components and their divisors are defined in relation to linear model (1).

Phenotypic and Genotypic Correlation

Phenotypic correlation (r_p) analysis was carried out using GenStat 7.1 (2003). The multivariate MANOVA procedure,

within GenStat 7.1 (2003), enabled estimation sums of cross-products, using the multisite trait data from the 40 HS families. Mean cross products were then calculated and resolved to estimate genotypic covariance components. The genotypic covariance components were used together with the σ_g^2 estimates, from REML analysis, to determine genotypic correlation coefficients (r_g) according to Falconer (1989).

Pattern Analysis

Pattern analysis was conducted to: (a) provide a graphical summary of the performance of the 40 HS families, six parental germplasm accessions and the two check cultivars of *M. officinalis*, based on the genotype \times trait BLUP adjusted mean matrix generated from variance component analysis across the two locations Yuzhong and Linze, and (b) investigate any changes in type (positive or negative) and magnitude of the association among the seven traits across Yuzhong and Linze. Pattern analysis consisted of a combination of cluster and principal component analysis (PCA) (Gabriel, 1971; Kroonenberg, 1994; Watson et al., 1995). To identify the optimum level of truncation for the resulting hierarchy from cluster analysis, the increase in the sum of squares among accession groups was monitored as the number of groups increased. The group level selected was determined by the point where the percentage of accession sum of squares among groups did not improve substantially as the number of groups increased (DeLacy, 1981).

RESULTS

Genotypic Variance Components and HS Family Mean Repeatability of Plant Attributes of *M. officinalis*

The genotypic variance estimated for the different traits from the individual location, Yuzhong and Linze, analysis indicated significant ($P < 0.05$) variation among the 40 *M. officinalis* HS families (Tables 2A,B). At both these locations HS family mean repeatability estimates ranged from intermediate to very high, depending on the traits.

At Yuzhong, the HS family mean repeatability (R_1) was high for the traits DW, SD and SV, which ranged from 0.89 to 0.96 (Table 2A). For the traits PH, LA and Cou, HS family mean repeatability was high (0.82–0.86). HS family mean repeatability was intermediate (0.60 and 0.70) for SN and LS. At Linze, HS family mean repeatability was very high (0.90–0.97) for the traits LA, SD, DW, and SV (Table 2B). The traits SY and SN had high (0.73, 0.77) HS family mean repeatability. HS family mean repeatability was intermediate (0.46, 0.53) for PH and LS.

Analysis of variance for mean trait expression across the two sites Yuzhong and Linze indicated significant ($P < 0.05$) genotypic variation among the 40 HS families. There was also significant ($P < 0.05$) genotype-by-location interaction, depending on the traits (Table 3). There was no significant ($P > 0.05$) genotype-by-location interaction for the traits LS and LA. Line mean repeatability (R_2) across the two locations varied from: relatively high for the traits DW and LA; intermediate for PH, LS, SD, and SV; and low for SN.

TABLE 2A | Average, maximum, minimum, least significant differences ($l.s.d_{0.05}$), genotypic (σ_g^2), and experimental error (σ_e^2) variance components and associated standard errors ($\pm SE$), and HS family mean repeatability (R_1) estimated from the 40 *M. officinalis* half sib families, evaluated at Yuzhong.

Traits	DW	LS	PH	SD	SN	SV	LA	Cou
Average	128	0.93	163	1.83	8.2	2.2	7.6	0.44
Max	216	1.22	192	2.77	11.4	4.7	8.3	1.14
Min	69	0.63	136	1.22	7.0	0.8	6.8	0.14
$l.s.d_{0.05}$	18	0.17	12	0.28	1.38	0.779	0.54	0.133
σ_g^2	1709 \pm 393	0.025 \pm 0.006	236 \pm 57	0.193 \pm 0.046	1.213 \pm 0.336	0.890 \pm 0.218	0.24 \pm 0.06	0.093 \pm 0.021
σ_e^2	232 \pm 17	0.032 \pm 0.002	153 \pm 11	0.069 \pm 0.005	2.43 \pm 0.18	0.208 \pm 0.030	0.114 \pm 0.017	0.004 \pm 0.001
R_1	0.96	0.70	0.82	0.89	0.60	0.93	0.86	0.86

LS, leaf to stem ratio; SR, spring vigor; LA, leaf area (cm^2); PH, plant height (cm); DW, herbage dry weight (g/plant); SD, stem diameter (cm); SN, stem number; Cou, coumarin (% of dry matter).

TABLE 2B | Average, maximum, minimum, least significant differences ($l.s.d_{0.05}$), genotypic (σ_g^2), and experimental error (σ_e^2) variance components and associated standard errors ($\pm SE$), and HS family mean repeatability (R_1) estimated from the 40 *M. officinalis* half sib families, evaluated at Linze.

Traits	DW	LS	PH	SD	SN	SV	LA	SY
Average	125	0.96	168	1.55	8.4	2.0	7.8	7.1
Max	207	1.18	190	2.63	11.5	3.5	8.5	10.3
Min	45	0.77	139	0.66	7.5	0.1	6.8	4.7
$l.s.d_{0.05}$	19	0.17	13	0.24	1.61	0.47	0.62	1.74
σ_g^2	1792 \pm 413	0.015 \pm 0.004	190 \pm 47.3	0.233 \pm 0.054	1.18 \pm 0.36	1.32 \pm 0.31	0.24 \pm 0.06	2.19 \pm 0.59
σ_e^2	264 \pm 19	0.040 \pm 0.003	170 \pm 12	0.047 \pm 0.003	3.80 \pm 0.27	0.11 \pm 0.02	0.08 \pm 0.01	2.48 \pm 0.23
R_1	0.95	0.53	0.46	0.94	0.77	0.97	0.90	0.73

LS, leaf to stem ratio; SV, spring vigor; LA, leaf area (cm^2); PH, plant height (cm); DW, herbage dry weight (g/plant); SD, stem diameter (cm); SN, stem number; SY, seed yield (g/plant).

TABLE 3 | Average, maximum, minimum, least significant differences ($l.s.d_{0.05}$), genotypic (σ_g^2), genotype-by-location interaction (σ_{gl}^2), and experimental error (σ_e^2) variance components and associated standard errors ($\pm SE$), and HS family mean repeatability (R_2) estimated from the 40 *M. officinalis* half sib families, evaluated across two locations, Yuzhong and Linze.

Traits	DW	LS	PH	SD	SN	LA	SV
Average	126	0.94	165	1.69	8.3	7.7	2.1
Max	206	1.20	188	2.57	11.6	8.4	3.8
Min	63	0.71	144	1.72	7.4	6.7	0.5
$l.s.d_{0.05}$	14	0.13	9	0.2	1.16	0.38	0.55
σ_g^2	1025 \pm 327	0.018 \pm 0.005	102 \pm 40	0.099 \pm 0.038	0.504 \pm 0.257	0.24 \pm 0.06	0.56 \pm 0.21
σ_{gl}^2	715 \pm 167	ns	109 \pm 28	0.112 \pm 0.027	0.637 \pm 0.227	ns	0.53 \pm 0.13
σ_e^2	279 \pm 14	0.036 \pm 0.002	167 \pm 9	0.065 \pm 0.003	3.52 \pm 0.18	0.09 \pm 0.01	0.18 \pm 0.02
R_2	0.72	0.60	0.55	0.60	0.36	0.89	0.65

LS, leaf to stem ratio; SV, spring vigor; LA, leaf area (cm^2); PH, plant height (cm); DW, herbage dry weight (g/plant); SD, stem diameter (cm); SN, stem number.
ns, not significant ($p < 0.05$).

Pattern Analysis: Principal Component Analysis (PCA)

The biplot (Figure 2) was generated from PCA of the 40 HS families, the six parental germplasm accessions and the two check cultivars of *M. officinalis*, based on the 9 traits LS, SV, LA, PH, DW, SD, SN, and Cou. The first principle component explained 46% of the total trait variation, and the second principle component explained 18%. The correlation structure of the traits is indicated by the directional vectors in the biplot. In this study,

SD, SN, and PH showed a strong positive association with DW. The traits LS and Cou also showed a negative correlation with DW.

The seven plant trait responses at the locations Yuzhong and Linze are presented in the two biplots, (Figures 3A,B). In Figure 3A, based on breeding line performance at Yuzhong, the first and second principal components accounted for 43 and 19% of the total variation, respectively. Based on the line performance at Linze, the first principle component explained 51% of the total

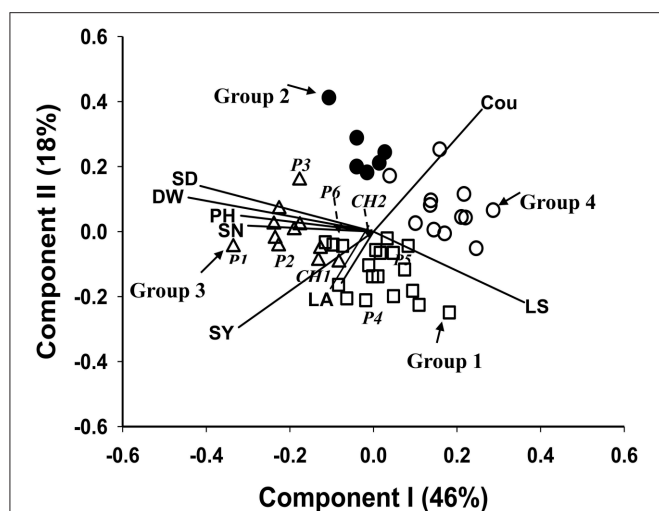


FIGURE 2 | Biplot generated using standardized Best Linear Unbiased Predictor values for eight traits measured from: the 40 half sib families, the 6 parental germplasm accessions and the 2 check cultivars of *M. officinalis*, evaluated across two locations Yuzhong and Linze.

Components I and II account for 46 and 18% of total variation, respectively. The different symbols indicate progeny Groups 1 to 4 generated from cluster analysis. The vectors represent the traits: LS, leaf to stem ratio; LA, leaf area; PH, plant height; DW, herbage dry weight; SD, stem diameter; SN, stem number; SY, seed yield; Cou, coumarin. The 6 parental germplasm accessions: P1 to P6. Check's: CH1, experimental cultivar; CH2, cv Norgold.

trait variation, and the second principle component explained 23% (Figure 3B).

There were differences in trait association across the two locations Yuzhong and Linze. The traits DW, SD, PH, SN, and LA showed a strong positive correlation at Yuzhong (angles between the directional vectors are at $<45^\circ$). At Linze, DW was positively correlated with SD and PH similar to that showed in Yuzhong. However, SN and LA showed a weak positive associated with DW (Figures 3A,B).

Cluster Analysis

Clustering of the 40 HS families, together with the 6 parental germplasm accessions and 2 check cultivars, was truncated at the four group level. Group 4, the largest group contained 17 members, followed by group 1, group 3 and group 2, which contained 14, 11, and 6 members, respectively (Table 4). As indicated by the Figure 2, the check cultivars were both in group 1. The parental germplasm accessions P1, P2, P3 and P4, P5, P6 were in groups 3 and 1, respectively. The trait means for each group (Table 4) indicated that the members in group 3 had high DW and low coumarin content, and those in group 1 had low coumarin content and intermediate expression for traits DW, PH, SD, and SN. The members in group 4 showed characteristics of a small plant type with high coumarin content. The highest expression for coumarin was in group 2. Groups 3 and 1 had higher SY expression in comparison to groups 4 and 2.

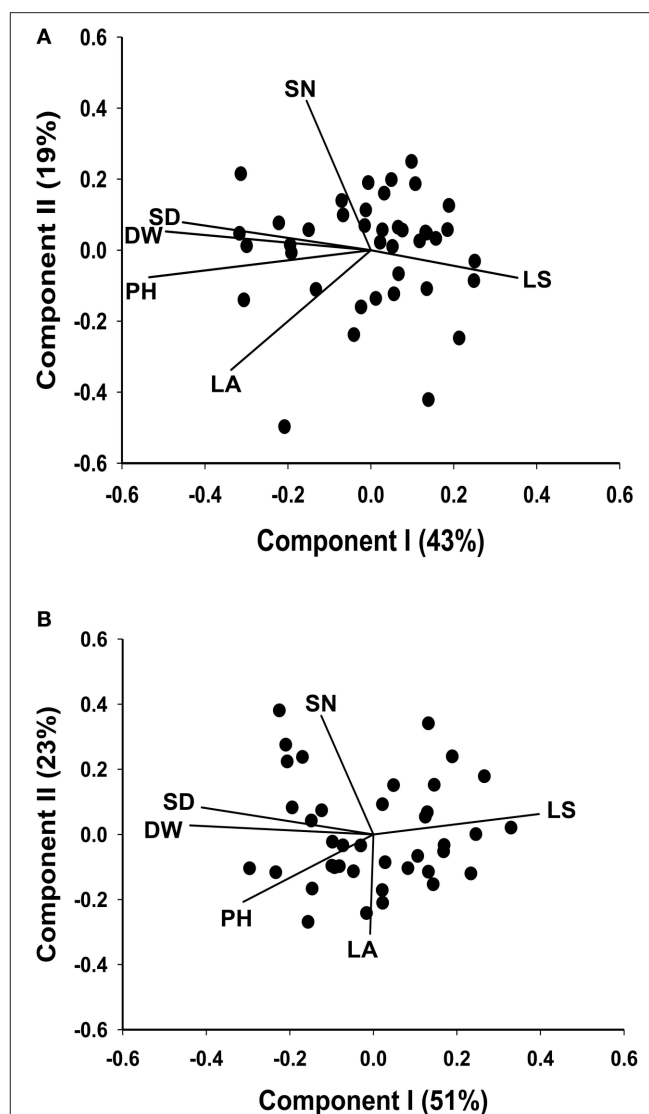


FIGURE 3 | Biplots based on standardized Best Linear Unbiased Predictor values for seven morphological traits, measured from the 40 half sib families of *M. officinalis*, evaluated at Yuzhong (A) and Linze (B). In each of the biplots Components I and II account for most of the total variation. The vectors represent the traits: LS, leaf to stem ratio; LA, leaf area; PH, plant height; DW, herbage dry weight; SD, stem diameter; SN, stem number.

Phenotypic and Genotypic Correlation

A range of genotypic and phenotypic correlation coefficients are presented in Table 5. These coefficients range from strong to weak positive and negative pairwise associations among the 7 traits. Of the special interest are the phenotypic and genotypic correlations between DW and the other traits. There was strong positive phenotypic correlation between DW and the traits SD, PH and SN, and strong negative phenotypic correlation with LS and SV. These results are further supported by the directional vectors in the biplots (Figures 2, 3A,B). In comparison to phenotypic correlation, the estimated genotypic

TABLE 4 | Trait means for each of the 4 half sib family groups generated from pattern analysis.

Group	No. members	DW	LS	PH	SD	SN	LA	SV	Cou	SY
1	14	104	0.995	162	1.45	7.50	7.74	2.52	0.365	6.07
2	6	123	0.849	164	1.80	7.75	7.58	2.41	1.986	1.99
3	11	157	0.905	168	2.00	8.73	7.87	1.65	0.405	6.54
4	17	73	1.125	150	1.20	7.49	7.58	3.47	1.561	3.00

LS, leaf to stem ratio; SR, spring vigor; LA, leaf area (cm²); PH, plant height (cm); DW, herbage dry weight (g/plant); SD, stem diameter (cm); SN, stem number; Cou, coumarin (% of dry matter); SY, seed yield (g/plant).

TABLE 5 | Genotypic (r_g) (lower triangle) and phenotypic (r_p) (upper triangle) correlation coefficients, between traits base on the 40 *M. officinalis* half sib families, the six parental germplasm accessions and the two check cultivars, evaluated across two locations, Yuzhong and Linze.

	DW	LA	L_S	PH	SD	SN	SV
DW		0.086	−0.371**	0.477**	0.658**	0.336**	−0.835**
LA	0.021		−0.023	0.116	0.166*	−0.070	−0.121
L_S	−0.129	−0.003		−0.227	−0.303	−0.037	0.330**
PH	0.082	0.037	−0.131		0.130*	0.127*	−0.510**
SD	0.051	0.011	−0.042	0.029		0.264**	−0.528**
SN	0.116	−0.045	−0.047	0.019	0.035		−0.361**
SR	−0.187	−0.029	0.139	−0.119	−0.057	−0.129	

LS, leaf to stem ratio; SV, spring vigor; LA, leaf area (cm²); PH, plant height (cm); DW, herbage dry weight (g/plant); SD, stem diameter (cm); SN, stem number. *, ** Significant at $p < 0.05$ and $p < 0.01$.

correlation coefficients for all 7 traits showed similar types of pairwise association (Table 5).

DISCUSSION

Previous studies on genotypic variation within *Melilotus* spp. have mainly focused on interspecific comparisons for traits such as coumarin content (Nair et al., 2010), salinity, waterlogging tolerance (Rogers et al., 2008), and also on phylogenetic relationships (Di et al., 2015) and genetic diversity (Di et al., 2014; Wu et al., 2016). The significant ($P < 0.05$) genotypic variation and high to moderate line mean repeatability reported from our study, indicates the potential for genetic improvement of the nine traits examined. There are no reported studies in *M. officinalis* similar to ours that estimate the magnitude of genotypic variation for key traits such as DW, Cou, PH, and SY.

Phenotypic variation, expressed as ranges, has been reported for some morphological traits. Klebesadel (1992) reported 2 year means of PH of *M. officinalis* ecotypes ranging from 112 to 145 cm. Second year mean plant height (PH) measured in our study ranged from 144 to 188 cm. Martino et al. (2006) reported a range of coumarin content between 0.12 and 0.39% based on different extraction methods. Nair et al. (2010) reported coumarin content measured from 27 *M. officinalis* accessions ranging from 0.09 to 0.61% of dry matter. Our study indicated a coumarin content that ranged from 0.04 to 0.91% of dry matter. Herbage dry matter from single plants has been reported from experiments conducted under glasshouse conditions (Rogers et al., 2008). There is a lack of information

on morphological traits measured under field conditions. Results from our study on the genotypic variation for the traits LS, SD, SN, LA, SV and SY, measured under field conditions, will be valuable to *Melilotus* breeders. Information on the magnitude and significance of the genotypic and environmental components of phenotypic variation for important traits will provide a basis for the development of efficient breeding methods for their improvement (Moll and Stuber, 1974). Results from the present study showed that there was significant genotypic variation among the 40 HS families at each location, Yuzhong and Linz, and also across these two locations for all the traits measured. High genotypic variation was present for DW, SV, and SD at Yuzhong and LA, SD, DW, and SV at Linze. These results, together with the relatively high HS family mean repeatabilities estimated, indicate the potential genetic variation available, within the new *M. officinalis* breeding population, for improvement of these traits through selection and breeding.

Forage plants are utilized across a wide range of environments, which include different climates, soil types and grazing systems (Breese, 1969). The presence of genotype-by-environment interactions complicates selection of material for broad adaptation due to unreliable performance across environments (Comstock and Moll, 1963; Cooper and Byth, 1996). Quantifying the magnitude and understanding the causes of genotype-by-environment interaction can be helpful when planning breeding strategies (Milligan et al., 1990; Basford and Cooper, 1998). Caradus (1993) reported that a range of traits in white clover, especially yield-related traits, were sensitive to genotype-by-environment interactions. A similar result in white clover was reported by Jahufer et al. (1999). In our study, the

genotype-by-environment interactions were significant for most traits except for the traits LS and LA. This indicates the importance of multi-site evaluation in *M. officinalis* breeding programs when focusing on broad adaptation. The application of multisite testing in breeding programs to investigate the effect of genotype-by-environment interaction on line performance has been reported for forage grass and legume species such as perennial ryegrass (Easton et al., 2015), switchgrass (Jahufer and Casler, 2015), alfalfa (Hill and Baylor, 1983), and white clover (Ballizany et al., 2012).

The association among the traits measured in our study was examined using a combination of phenotypic and genotypic correlation with pattern analysis. The estimates of phenotypic and genotypic correlation coefficients supported the association among traits indicated in the biplots. The positive and significant phenotypic association of DW with traits PH, SD, and SN, predicts a positive correlated response in all these traits when any one of them is selected for individual. This relationship will be useful in a breeding program. The strong positive correlation between DW and SY shown in the biplot (Figure 2) indicates that selection for herbage yield would also result in increasing seed yield. Significant correlation of forage yield and seed yield was also demonstrated in other legumes (Iannucci and Martiniello, 1998; Guler et al., 2001; Cakmakci et al., 2006). Our study indicated negative phenotypic and genotypic correlation between DW and LS. The LS is used as an indicator of digestibility and intake in forage (Kephart et al., 1990). This result implies a trade-off between herbage yield and quality. Julier et al. (2000) also estimated significant negative correlation between DW and LS in alfalfa, which is similar to *M. officinalis* in vegetative form (Whitson et al., 1992).

The strong negative relationship between SV and DW suggests that measurement of spring vigor, at a very early stage of plant growth, could serve as an indirect selection criterion for increasing herbage yield for *M. officinalis* grown in western China (Table 5). This will increase the efficiency of current breeding methods, especially when dealing with the biennial forage species like *M. officinalis*. Similar results were reported from studies on common vetch (Cakmakci et al., 2006). The negative phenotypic correlation between the traits DW and Cou shown in our study (Figure 2) indicates the possibility of identifying HS families with a combination of high herbage dry weight and low coumarin content expression. This association will be of significant importance in our *M. officinalis* breeding program. Hofmann and Jahufer (2011) showed negative association between flavonoid accumulation and biomass using multivariate analysis.

Pattern analysis has been successfully used to summarize complex genotype-by-environment (Cooper et al., 1993; Zhang et al., 2006) and genotype-by-trait (Jahufer et al., 1999; Davodi et al., 2011) data matrices. Jahufer et al. (1999) successfully

identified superior white clover full-sib families based on seven morphological traits using a combination of principle component and cluster analysis. Davodi et al. (2011) used pattern analysis to summarize the performance of 200 alfalfa germplasm accessions, based on 12 traits, for use in the improvement of yield and quality. In our study, pattern analysis generated four groups (Figure 3), where group 3 consisted of HS families with above average performance for DW and below average performance for Cou. Group 3 consisted of 11 members, which included the parental germplasm accessions P1, P2, and P3. All the HS families in group 3 had a higher expression of the traits DW, SD, and SY in comparison to both commercial checks. The breeding lines in group 3 will be polycrossed to produce a breeding population that will be used in the recurrent selection program to develop new cultivars of *M. officinalis* with high herbage yield and low coumarin content for the Loess Plateau region in China.

CONCLUSION

The estimates of genotypic variation and HS family mean repeatability indicate the potential genetic variation available for all the traits examined in our study. These estimates also indicate the potential to develop cultivars with increased forage yield and low coumarin content. The significant genotype-by-environment interaction estimated for the traits DW, PH, SD, SN, and SV across the two environments, Yuzhong and Linze, indicate the importance of multi-environment evaluation trials in our *M. officinalis* breeding program. The breeding population developed by polycrossing the HS families within group 3, identified using pattern analysis, will provide a significant breeding pool for *M. officinalis* cultivar development in China.

AUTHOR CONTRIBUTIONS

KL, MJ, JZ, and YW conceived the topic. KL, FW, HD, and XM performed the experiments. KL and MJ analyzed all statistical data. KL wrote the manuscript. All authors revised the manuscript. We thank National Plant Germplasm System (NPGS) for offering the *Melilotus officinalis* seeds.

ACKNOWLEDGMENTS

This work was supported by National Basic Research Program (973) of China (2014CB138704), Special Fund for Agroscientific Research in the Public Interest (20120304205), Natural Science Foundation of China (31572453), Program for Changjiang Scholars and Innovative Research Team in University (IRT13019), and Fundamental Research Funds for the Central Universities (lzujbky-2016-10).

REFERENCES

- Allen, O. N., and Allen, E. K. (1981). *The Leguminosae: A Source Book of Characteristics, Uses and Nodulation*. Madison, WI: The University of Wisconsin Press.
- Ballizany, W. L., Hofmann, R. W., Jahufer, M. Z. Z., and Barrett, B. A. (2012). Genotype \times environment analysis of flavonoid accumulation and morphology in white clover under contrasting field conditions. *Field Crop. Res.* 128, 156–166. doi: 10.1016/j.fcr.2011.12.006

- Basford, K. E., and Cooper, M. (1998). Genotype \times environment interactions and some considerations of their implications for wheat breeding in Australia. *Aust. J. Agr. Res.* 49, 153–174. doi: 10.1071/A97035
- Breese, E. L. (1969). The measurement and significance of genotype – environment interactions in grasses. *Heredity* 24, 27–44. doi: 10.1038/hdy.1969.3
- Breese, E. L., and Hayward, M. D. (1972). The genetic basis of present breeding methods in forage crops. *Euphytica* 21, 324–336.
- Cakmakci, S., Aydinoglu, B., Karaca, M., and Bilgen, M. (2006). Heritability of yield components in common vetch (*Vicia sativa* L.). *Acta Agric. Scand. B Soil Plant.* 56, 54–59. doi: 10.1080/09064710510008531
- Caradus, J. R. (1993). “White clover breeding line performance under sheep and cattle grazing,” in *Proceedings of the 10th Australian Plant Breeding Conference*, Vol. 2 (Gold Coast, QLD), 35–36.
- Comstock, R. E., and Moll, R. H. (1963). “Genotype-environment interactions,” in *Statistical Genetics and Plant Breeding*, Vol. 982, eds W. D. Hanson and H. F. Robinson (Washington, DC: Nat Acad Sci Natl Res Publ), 164–196.
- Cooper, M., and Byth, D. E. (1996). “Understanding plant adaptation to achieve systematic applied crop improvement - A fundamental challenge,” in *Plant Adaptation and Crop Improvement*, eds M. Cooper and G. L. Hammer (Wallingford, CT: CAB International), 5–23.
- Cooper, M., Delacy, I. H., Byth, D. E., and Woodruff, D. R. (1993a). Predicting grain-yield in Australian environments using data from CIMMYT international wheat performance trials. 2. The application of classification to identify environmental relationships which exploit correlated response to selection. *Field Crop. Res.* 32, 323–342.
- Cooper, M., DeLacy, I. H., and Eismann, R. L. (1993b). “Recent advances in the study of genotype \times environment interactions and their application to plant breeding,” in *Proceedings of the 10th Australian Plant Breeding Conference*, Vol. 1 (Gold Coast, QLD), 16–31.
- Davodi, M., Jafari, A. A., Assadian, G., and Ariapour, A. (2011). Assessment of relationships among yield and quality traits in alfalfa (*Medicago sativa*) under dryland farming system, Hamadan, Iran. *J. Rangeland Sci.* 1, 1247–1254.
- DeLacy, I. H. (1981). “Cluster analysis for the interpretation of genotype by environment interaction,” in *Interpretation of Plant Response and Adaptation to Agricultural Environments*, eds D. E. Byth and V. E. Mungomery (Brisbane, QLD: Australian Institute of Agricultural Science), 277–292.
- Di, H. Y., Duan, Z., Luo, K., Zhang, D. Y., Wu, F., Zhang, J. Y., et al. (2015). Interspecific phylogenetic relationships within genus *Melilotus* based on nuclear and chloroplast DNA. *PLoS ONE* 10:e0132596. doi: 10.1371/journal.pone.0132596
- Di, H. Y., Luo, K., Zhang, J. Y., Duan, Z., Huo, Y. X., and Wang, Y. R. (2014). Genetic diversity analysis of *Melilotus* populations based on ITS and *trnL-trnF* Sequences. *Acta Bot. Boreali Occident. Sin.* 34, 0265–0269. doi: 10.7606/j.issn.1000-4025.2014.02.0265
- Easton, H. S., Jahufer, M. Z. Z., Flay, C., Schmidt, J., Rolston, M. P., Trethewey, J. A. K., et al. (2015). Environment and ryegrass genetics – multisite trial of advanced breeding populations. *J. N.Z. Grasslands* 77, 233–238.
- Evans, P. M., and Kearney, G. A. (2003). *Melilotus albus* (Medik) is productive and persistent on saline soils of neutral to alkaline reaction in the high rainfall zone of south-west Victoria. *Aust. J. Exp. Agric.* 43, 349–355. doi: 10.1071/EA02079
- Falconer, D. S. (1989). *Introduction to Quantitative Genetics*. New York, NY: Longman Scientific and Technical.
- Fehr, W. R. (1987). *Principles of Cultivar Development*, Vol. 1. New York, NY: Collier Macmillan Publishers.
- Gabriel, K. R. (1971). The biplot graphical display of matrices with application to principal component analysis. *Biometrika* 58, 453–467. doi: 10.1093/biomet/58.3.453
- Goplen, B. P. (1971). Polara, a low coumarin cultivar of sweetclover. *Can. J. Plant Sci.* 51, 249–251. doi: 10.4141/cjps71-050
- Goplen, B. P. (1981). NORGOLD - A low coumarin yellow blossom sweetclover. *Can. J. Plant Sci.* 61, 1019–1021. doi: 10.4141/cjps81-154
- Gorz, H. J., Haskins, F. A., Manglitz, G. R., Smith, R. R., and Vogel, K. P. (1992). Registration of N28 and N29 sweetclover germplasms. *Crop Sci.* 32, 510.
- GRIN (2000). *Grin Taxonomy*. United States Department of Agriculture, Agricultural Research Service. Baltimore, MD: The Germplasm Resources Information Network (GRIN).
- Guler, M., Adak, M. S., and Ulukan, H. (2001). Determining relationships among yield and some yield components using path coefficient analysis in chickpea (*Cicer arietinum* L.). *Eur. J. Agric.* 14, 161–166. doi: 10.1016/S1161-0301(00)00086-1
- Hallauer, A. R., and Miranda, J. B. (1981). *Quantitative Genetics in Maize Breeding*. Ames: IA State University Press.
- Hill, R. R., and Baylor, J. E. (1983). Genotype \times environment interaction analysis for yield in alfalfa. *Crop Sci.* 23, 811–815. doi: 10.2135/cropsci1983.0011183X002300050002x
- Hofmann, R. W., and Jahufer, M. Z. Z. (2011). Tradeoff between biomass and flavonoid accumulation in white clover reflects contrasting plant strategies. *PLoS ONE* 6:e18949. doi: 10.1371/journal.pone.0018949
- Hu, B., Yu, J., Zhao, Z. H., Li, F. M., and Siddique, K. H. M. (2012). Soil P availability, inorganic P fractions and yield effect in a calcareous soil with plastic-film-mulched spring wheat. *Field Crop. Res.* 137, 221–229. doi: 10.1016/j.fcr.2012.08.014
- Iannucci, A., and Martiniello, P. (1998). Analysis of seed yield components in four mediterranean annual clovers. *Field Crop. Res.* 55, 235–243. doi: 10.1016/S0378-4290(97)00091-9
- Ivanov, A. I., and Chetvertnykh, L. M. (1980). Evaluation of a collection of *Melilotus* species and varieties, and variation in their characters. *Vestn. Selskokhozyaistvennoi Nauki* 5, 117–122.
- Jahufer, M. Z. Z., and Casler, M. (2015). Application of the smith-hazel selection index for improving biomass yield and quality of switchgrass. *Crop Sci.* 55, 1212–1222. doi: 10.2135/cropsci2014.08.0575
- Jahufer, M. Z. Z., and Cooper, M., Ayres, J. F., and Bray, R. A. (2002). Identification of future research to improve the efficiency of conventional white clover breeding strategies in Australia – a review. *Aust. J. Agric. Res.* 53, 239–256. doi: 10.1071/AR01110
- Jahufer, M. Z. Z., Cooper, M., Bray, R. A., and Ayres, J. F. (1999). Evaluation of white clover (*Trifolium repens* L.) populations for summer moisture stress adaptation in Australia. *Aust. J. Agric. Res.* 50, 561–574. doi: 10.1071/A98141
- Julier, B., Christian, H., and Christian, E. (2000). Within- and among-cultivar genetic variation in alfalfa: forage quality, morphology, and yield. *Crop Sci.* 40, 365–369. doi: 10.2135/cropsci2000.402365x
- Kephart, K. D., Buxton, D. R., and Hill, R. R. (1990). Digestibility and cell-wall components of alfalfa following selection for divergent herbage lignin concentration. *Crop Sci.* 30, 207–212. doi: 10.2135/cropsci1990.0011183X003000010045x
- Klebesadel, L. J. (1992). *Morphological, Physiological, and Winterhardness Comparisons among Latitudinal Ecotypes of Biennial Sweetclover (Melilotus Species) in Subarctic Alaska*. Fairbanks, AK: School of Agriculture and Land Resources Management Agricultural and Forestry Experiment Station.
- Kroonenberg, P. M. K. (1994). The TUCKALS line: a suite of programs for three-way data analysis. *Comput. Stat. Data Anal.* 18, 73–96. doi: 10.1016/0167-9473(94)90133-3
- Li, X. Y., Wang, Y. R., Wei, X., Tai, J. H., Jia, C. Z., Hu, X. W., et al. (2014). Planting density and irrigation timing affects *Cleistogenes songorica* seed yield sustainability. *Agron. J.* 106, 1690–1696. doi: 10.2134/agronj14.0111
- Luo, K., Di, H.-Y., Zhang, J.-Y., Wang, Y.-R., and Li, Z.-Q. (2014). Preliminary evaluation of agronomy and quality traits of nineteen *Melilotus* accessions. *Pratacultural Sci.* 31, 2125–2134. doi: 10.11829/j.issn.1001-0629.2013-0455
- Maddaloni, J. (1986). Forage production on saline and alkaline soils in the humid region of Argentina. *Reclam. Revag. Res.* 5, 11–16.
- Martino, E., Ramaiola, I., Urbano, M., Bracco, F., and Collina, S. (2006). Microwave-assisted extraction of coumarin and related compounds from *Melilotus officinalis* (L.) pallas as an alternative to soxhlet and ultrasound-assisted extraction. *J. Chromatogr. A.* 1125, 147–151. doi: 10.1016/j.chroma.2006.05.032
- Milligan, S. B., Gravois, K. A., Bischoff, K. P., and Martin, F. A. (1990). Crop effects on broad-sense heritabilities and genetic variances of sugarcane yield components. *Crop Sci.* 30, 344–349. doi: 10.2135/cropsci1990.0011183X003000020020x
- Moll, R. H., and Stuber, C. W. (1974). Quantitative genetics-empirical results relevant to plant breeding. *Advn. Agron.* 26, 277–313. doi: 10.1016/S0065-2113(08)60874-3
- Nair, R. M., Whittall, A., Hughes, S. J., Craig, A. D., Revell, D. K., Miller, S. M., et al. (2010). Variation in coumarin content of *Melilotus* species grown in South Australia. *N.Z. J. Agr. Res.* 53, 201–213. doi: 10.1080/00288233.2010.495743

- Piano, E., Annicchiarico, P., Romani, M., and Pecetti, L. (2007). Effect of the number of parents and their combining ability on the performance of synthetic varieties in tall fescue. *Aust. J. Agr. Res.* 58, 1100–1105. doi: 10.1071/AR06424
- Riday, H., and Brummer, C. (2007). Narrow sense heritability and additive genetic correlations in alfalfa subsp. *falcata*. *J. Iowa Acad. Sci.* 114, 28–34.
- Rogers, M. E., Colmer, T. D., Frost, K., Henry, D., Cornwall, D., Hulm, E., et al. (2008). Diversity in the genus *Melilotus* for tolerance to salinity and waterlogging. *Plant Soil* 304, 89–101. doi: 10.1007/s11104-007-9523-y
- Sagalbekov, U. M. (1980). Evaluation of a collection of sweet clover in the southern forest steppe of omsk province. *Nauch. Tekhn. Byul. Sib. NII S. Kh.* 3–6.
- Sherif, E. A. (2009). *Melilotus indicus* (L.) all, a salt-tolerant wild leguminous herb with high potential for use as a forage crop in salt-affected soils. *Flora* 204, 737–746. doi: 10.1016/j.flora.2008.10.004
- Smith, W. K., and Gorz, H. J. (1965). Sweetclover improvement. *Adv. Agron.* 17, 163–231. doi: 10.1016/S0065-2113(08)60414-9
- Su, Y. Z., Zhao, W. Z., Su, P. X., Zhang, Z. H., Wang, T., and Ram, R. (2007). Ecological effects of desertification control and desertified land reclamation in an oasis-desert ecotone in an arid region: a case study in Hexi Corridor, northwest China. *Ecol. Eng.* 29, 117–124. doi: 10.1016/j.ecoleng.2005.10.015
- Turkington, R. A., Cavers, P. B., and Rempel, E. (1978). The biology of Canadian weeds. 29. *Melilotus alba* Desr. and *M. officinalis* (L.) Lam. *Can. J. Plant Sci.* 58, 523–537. doi: 10.4141/cjps78-078
- Watson, S. L., DeLacy, I. H., Podlich, D. W., and Basford, K. E. (1995). *GEBEIL: An Analysis Package Using Agglomerative Hierarchical Classificatory and Svd Ordination Procedures for Genotype × Environment Data*. Brisbane, QLA: Centre for Statistics Research Report; Department of Agriculture, University of Queensland.
- White, T. L., and Hodge, G. R. (1989). “Predicting breeding values with applications in forest tree improvement,” in *Forestry Sciences* 33, ed P. K. R. Nair (Boston, MA: Kluwer Academic), 367.
- Whitson, T. D., Burrill, L. C., Dewey, S. A., Cudney, D. W., Nelson, B. E., Lee, R. D., et al. (1992). *Weeds of the West* (Edited by T. D. Whitson). Laramie: The Western Society of Weed Science in cooperation with the Western United States Land Grant Universities Cooperative Extension Services and the University of Wyoming.
- Wu, F., Zhang, D. Y., Ma, J. X., Luo, K., Di, H. Y., Liu, Z. P., et al. (2016). Analysis of genetic diversity and population structure in accessions of the genus *Melilotus*. *Ind. Crop. Prod.* 85, 84–92. doi: 10.1016/j.indcrop.2016.02.055
- Zhang, Y., He, Z. H., Zhang, A. M., Ginkel, M. V., Pena, R. J., and Ye, G. Y. (2006). Pattern analysis on protein properties of Chinese and CIMMYT spring wheat cultivars sown in China and CIMMYT. *Aust. J. Agric. Res.* 57, 811–822. doi: 10.1071/AR05372
- Zhu, H., and Fan, J. (2008). HPLC determination of coumarin in *Melilotus*. *Chin. J. Pharm. Anal.* 28, 2111–2113. doi: 10.16155/j.0254-1793.2008.12.024

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Luo, Jahufer, Wu, Di, Zhang, Meng, Zhang and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Multidisciplinary Contributions to Legume Crop History: Proceed with Caution

Frank M. Dugan *

USDA-ARS Plant Introduction, Washington State University, Pullman, WA, USA

Keywords: archeobotany, crop history, paleolinguistics, pulses, legumes

OPEN ACCESS

Edited by:

Nicolas Rispaïl,
Spanish National Research Council
(CSIC), Spain

Reviewed by:

Guus Kroonen,
Leiden University, Netherlands
Kevin E. McPhee,
North Dakota State University, USA

*Correspondence:

Frank M. Dugan
frank.dugan@ars.usda.gov

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 18 July 2016

Accepted: 28 November 2016

Published: 20 December 2016

Citation:

Dugan FM (2016) Multidisciplinary
Contributions to Legume Crop
History: Proceed with Caution.
Front. Plant Sci. 7:1876.
doi: 10.3389/fpls.2016.01876

Accurate understanding of crop biogeography facilitates comprehension of agronomic potential, genetic diversity, and crop–pathogen evolution. Classic perspectives are exemplified by Vavilov (1987), a posthumous compilation, and Harlan (1971). Origin and geographic spread of a given crop provide clues as to environmental interactions, including relative adaption to pests, pathogens, and abiotic factors (Dark and Gent, 2001; Dugan, 2015). Great antiquity of pulse crops (pea, chickpea, lentil, bitter vetch, faba bean) is documented archeobotanically for the Fertile Crescent and several adjacent areas, including much of Europe in succeeding times (e.g., Abbo et al., 2003, 2006; Zohary et al., 2012; Mikić et al., 2014).

These crops are also indicated in ancient texts of the Fertile Crescent and adjacent areas, e.g., Akkadian, Old Babylonian (Semitic languages), Hittite (Indo-European, IE), and Sumerian (reviewed in Dugan, 2015). Linguistic indications of pulse crops in ancient Greek, Latin, Old Slavic, etc., and more recent languages have been summarized (Mikić, 2012). However, there remain areas in which data are scarce or in which there is conspicuous dissent from consensus. This is especially true for the Bronze Age Steppes, now seen as critically important in dissemination of language, culture, and peoples (Haak et al., 2015), and possibly the geographic location for the ancestral Proto-Indo-Europeans (Anthony, 2007).

Archeobotanical data for the region remain suboptimal due to the infrequency of applying modern techniques for recovery and dating of plant remains (Mallory, 2014), but repeated instances of cereal chaff imprints or chaff itself on pottery (~4500 to 3500 BC) imply cropping of cereals (Motuzaite-Matuzeviciute, 2012). Presence of pulse crops is not yet conclusively documented, and they may have been absent, although they are present in the archeobotanical record for regions bordering the Steppes (Dugan, 2015). Archeobotanical evidence for pulses appears lacking in the Steppes when archeobotanical sites with recovery of pulses are mapped (Mikić, 2012, 2015a, 2016; Mikić et al., 2014). One possible explanation for archeological recovery of remains of cereals in the Steppes, but not of pulses, is that Bronze Age Steppes were semiarid, as indicated by paleoclimatic studies (Alekseeva et al., 2007; Khomutova et al., 2007; Mitusov et al., 2009). Estimates of mean annual precipitation imply climate may have been unsuitable for consistent cultivation of legumes, but typically with sufficient moisture for cereals (Dugan, 2015).

We have no literature in Proto-Indo-European (PIE, a reconstructed language), analogous to the agricultural writings of ancient civilizations. But perhaps analyses of PIE can promote understanding of Bronze Age agriculture with regard to pulses. Initial inspection of agronomic literature gives reason to believe so. Mikić (2016), indicates a putatively PIE root **erag^w[h]* – (“a kernel of a leguminous plant”) as originating in the Steppes just northeast of the Black Sea. Mikić (2015a,b) attributes to PIE additional words for legumes, including pea and lentil, as do Mikić (2009, 2011, 2012) and Mikić et al. (2008). The notion that PIE contains words for legumes is now embedded in agronomic literature, including this journal (Mikić, 2015c). Cited in justification (e.g., in Mikić, 2009, 2012) are Vasmer (1953), the slightly more recent Pokorný (1959, 1969) and a

website, The Tower of Babel¹. The website references “Fraenkel 358” for PIE **lent-* (lentil), and also “Pokorny’s dictionary.” Fraenkel (1962) contains “lent” on page 358, but not as PIE and not lentil. I could not locate the putative PIE terms from Mikić (2009, 2012) in Vasmer (1953), although Vasmer provides Slavic roots similar to these terms. Also, “Pokorny (1959) is badly out of date; moreover it errors extravagantly on the side of inclusion, listing every word ... that might conceivably reflect a PIE lexeme” (Ringe, 2006). It therefore becomes necessary to inspect literature from contemporary linguists specifically addressing PIE, and in doing so a different result is obtained. We find that words for cereals were abundant in PIE (Mallory and Adams, 1997, 2006), but the case is more ambiguous for pulses.

At present, the consensus seems to be that words for pulses appear in later regional terms (Indo-European, but not PIE) or represent borrowings (“loan words”) from non-Indo-European languages. Mallory and Adams (1997) note with regard to pea, “our inability to reconstruct in a solid way any PIE word for it.” These authors state for vetch (*Vicia sativa*), bitter vetch (*Vicia ervila*), and grass pea (*Lathyrus sativus*), “there is no evidence for IE antiquity [i.e., PIE] for any of these crops [and] the lentil (*Lens culinaris*) is also unretrievable from the IE lexicon.” Mallory and Adams (2006) state with regard to pea and chickpea, “their designations are found only regionally ... (Latin, Greek), which raises the possibility that they may derive from a non-IE substratum.” Somewhat more ambiguous is a possible PIE reconstruction for bean, **bhabheh*_a-, although Mallory and Adams (2006) indicate this word as among “regional terms” eventually yielding Greek, Latin, and Albanian equivalents. That names of pulses in Greek, Latin, and some other IE languages originate from a Pre-Greek, non-Indo-European substratum has long been the opinion (Sturtevant, 1910; Mann, 1943; Hester, 1968), and this consensus has held (e.g., Adrados, 2005; Kroonen, 2012; Darden, 2013). This does not mean that Proto-Indo-Europeans were utterly unfamiliar with pulses, but consensus among those just cited is that words specifically designating pulses are *not* found in reconstructed PIE.

¹<http://starling.rinet.ru/>.

REFERENCES

- Abbo, S., Frenkel, O., Sherman, A., and Shtienberg, D. (2006). The sympatric Ascochyta pathosystems of Near Eastern legumes, a key for better understanding of pathogen biology. *Eur. J. Plant Pathol.* 119, 111–118. doi: 10.1007/s10658-007-9116-x
- Abbo, S., Shtienberg, D., Lichtenzieg, J., Lev-Yadun, S., and Gopher, A. (2003). The chickpea, summer cropping, and a new model for pulse domestication in the ancient Near East. *Q. Rev. Biol.* 78, 435–448. doi: 10.1086/378927
- Adrados, F. R. (2005). *A History of the Greek Language*. Leiden: Brill.
- Alekseeva, T., Alexseev, A., Maher, B. A., and Demkin, V. (2007). Late Holocene climate reconstructions for the Russian steppe, based on mineralogical and magnetic properties of buried palaeosols. *Palaeogeogr. Paleoclim. Paleoecol.* 249, 103–127. doi: 10.1016/j.palaeo.2007.01.006
- Anthony, D. (2007). *The Horse, the Wheel, and Language: How Bronze Age Riders from the Eurasian Steppes Shaped the Modern World*. Princeton, NJ: Princeton University Press.
- Bouckaert, R., Lemey, P., Dunn, M., Greenhill, S. J., Alekseyenko, A. V., Drummond, A. J., et al. (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337, 957–960. doi: 10.1126/science.1219669
- Darden, B. J. (2013). “The satem shift, Armenian isern, and the early Indo-European of the Balkans,” in *Language Typology and Historical Contingency, Typological Studies in Language*, Vol. 104, eds B. Bickel, L. A. Grenoble, D. A. Peterson, and A. Timberlake (Amsterdam: John Benjamins BV), 229–308.
- Dark, P., and Gent, H. (2001). Pests and diseases of prehistoric crops: a yield “honeymoon” for early grain crops in Europe? *Oxford J. Archaeol.* 20, 59–78. doi: 10.1111/1468-0092.00123
- Dugan, F. M. (2015). Shadows of Works and Days: “A revised history of cereals, legumes, and their pathogens from Neolithic through premodern Europe,” in *Hidden Histories and Ancient Mysteries of Witches, Plants, and Fungi* (St. Paul, MN: APS Press), 1–36.
- Fraenkel, E. (1962). *Litauisches Etymologisches Wörterbuch*, Vol. 1. Heidelberg: C. Winter.
- Origins of the Proto-Indo-Europeans and PIE have long been sought in Anatolia (Renfrew, 1990; Bouckaert et al., 2012) or, in a competing hypothesis, the Steppes of Ukraine and Russia (Anthony, 2007). The Steppes and Anatolian hypotheses are relevant to historical biogeography of agricultural crops. Absence, or near absence, of PIE words denoting pulses, and the lack, or near lack, of archeobotanical records for pulses in the Bronze Age Steppes, would correlate with paleoclimatic data indicating that the Steppes of the Bronze Age were generally too arid for successful cultivation of pulses. Both the Steppes and Anatolia provided adequate conditions for grazing livestock and for cereals (Hald, 2010; Motuzaite-Matuzeviciute, 2012), all of which are repeatedly manifest in PIE. Anatolia, however, also provided conditions adequate for pulses, and pulses are substantially present in the archeobotanical record there (e.g., Sadori et al., 2006). If the origins of PIE are sought in Anatolia, the consistent lack of PIE words for pulses, as noted by Mallory, Adams and their colleagues, requires explanation. If Mikić and his colleagues are correct, no such explanation is necessary under the Anatolian hypothesis.
- An accurate understanding of pulses in Proto-Indo-European and Indo-European agriculture and language will have positive consequences for our understanding of archeology, linguistics, and crop biogeography. A compilation of PIE agricultural terms is forthcoming in proceedings from a conference in Leipzig (J. Mallory and G. Kroonen, editors; J. Mallory, personal communication).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

ACKNOWLEDGMENTS

The author thanks Shahal Abbo and George Vandemark for comments on the manuscript, and James Mallory, David Anthony, and Aleksandar Mikić for instructive correspondence.

- Haak, W., Lazaridis, I., Patterson, N., Rohland, N., Mallick, S., Llamas, B., et al. (2015). Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature* 522, 207–211. doi: 10.1038/nature14317
- Hald, M. M. (2010). Distribution of crops at late Early Bronze Age Titriş Höyük, southeast Anatolia: towards a model for the identification of consumers of centrally organised food distribution. *Veget. Hist. Archaeobot.* 19, 69–77. doi: 10.1007/s00334-009-0223-9
- Harlan, J. R. (1971). Agricultural origins: centers and noncenters. *Science* 174, 468–474. doi: 10.1126/science.174.4008.468
- Hester, D. A. (1968). Recent developments in Mediterranean “substrate” studies. *Minos* 9, 219–235.
- Khomutova, T. E., Demkina, T. S., Borisov, A. V., Kashinskaya, N. N., Yeltsov, M. V., and Demkin, V. A. (2007). An assessment of changes in properties of steppe kurgan paleosols in relation to prevailing climates over recent millennia. *Quat. Res.* 67, 328–336. doi: 10.1016/j.yqres.2007.01.001
- Kroonen, G. (2012). Non-Indo-European root nouns in Germanic: evidence in support of the Agricultural Substrate Hypothesis. *Mém. Soc. Finno Ougrienne* 266, 239–260. Available online at: http://www.sgr.fi/sust/sust266/sust266_kroonen.pdf
- Mallory, J. P. (2014). “Indo-European dispersals and the Eurasian Steppe,” in *Reconfiguring the Silk Road: New Research on East-West Exchange in Antiquity*, eds V. H. Mair and J. Hickman (Philadelphia, PA: University of Pennsylvania Press), 73–88.
- Mallory, J. P., and Adams, D. Q. (1997). *Encyclopedia of Indo-European Culture*. Chicago, IL: Fitzroy Dearborne.
- Mallory, J. P., and Adams, D. Q. (2006). *The Oxford Introduction to Proto-Indo-European and the Proto-Indo-European World*. Oxford: Oxford University Press.
- Mann, S. E. (1943). The cradle of the Indo-Europeans: linguistic evidence. *Man* 43, 74–85. doi: 10.2307/2792846
- Mikić, A. (2009). Words denoting pea (*Pisum sativum*) in European languages. *Pisum Genet.* 41, 29–33. Available online at: <http://hermes.bionet.nsc.ru/pg/41/29.pdf>
- Mikić, A. (2011). Words denoting faba bean (*Vicia faba*) in European languages. *Field Vegetable Crop Res.* 48, 233–238. Available online at: <http://scindeks-clanci.ceon.rs/data/pdf/1821-3944/2011/1821-39441101233M.pdf>
- Mikić, A. (2012). Origin of the words denoting some of the most ancient Old World pulse crops and their diversity in modern European languages. *PLoS ONE* 7:e44512. doi: 10.1371/journal.pone.0044512
- Mikić, A. (2015a). Paleolinguistics and ancient Eurasian pulse crops. *Curr. Sci.* 108, 45–50.
- Mikić, A. (2015b). Root words relating to grain legumes in the proto-languages of Asia, Europe and North Africa. *Legume Perspect.* 6, 7–8. Available online at: <http://ils.nsseme.com/assets/LegumPerspect6.pdf>
- Mikić, A. (2015c). The first attested extraction of ancient DNA in legumes (Fabaceae). *Front. Plant Sci.* 6:1006. doi: 10.3389/fpls.2015.01006
- Mikić, A. (2016). Presence of vetches (*Vicia* spp.) in agricultural and wild floras of ancient Europe. *Genet. Resour. Crop Evol.* 63, 745–754. doi: 10.1007/s10722-016-0382-3
- Mikić, A., Čupina, B., Mihailović, V., Vasić, M., and Đorđević, V. (2008). “Diversity of words denoting traditional annual legumes in modern European languages,” in *Modern Variety Breeding for Present and Future Needs, Proceedings of the 18th EUCARPIA General Congress Valencia, Spain, 9–12 Sept 2008*, eds J. Prohens and M. B. Badenes (Valencia: Universidad Politécnica de Valencia), 193–197.
- Mikić, A., Medović, A., Javanović, Ž., and Stanisavljević, N. (2014). Integrating archaeobotany, paleogenetics and historical linguistics may cast more light onto crop domestication: the case of pea (*Pisum sativum*). *Genet. Resour. Crop. Evol.* 61, 887–892. doi: 10.1007/s10722-014-0102-9
- Mitusov, A. V., Mitusova, O. E., Pustovoytov, K., Lubos, C. C.-M., Dreibrodt, S., and Bork, H.-R. (2009). Paleoclimatic indicators in soils buried under archaeological monuments in the Eurasian steppe: a review. *Holocene* 19, 1153–1160. doi: 10.1177/0959683609345076
- Motuzaitė-Matuzevičiute, G. (2012). The earliest appearance of domesticated plant species and their origins on the western fringes of the Eurasian Steppe. *Documenta Praehistorica* 39, 1–21. doi: 10.4312/dp.39.1
- Pokorny, J. (1959). *Indogermanisches Etymologisches Wörterbuch*. Vol. 1. Bern: Francke.
- Pokorny, J. (1969). *Indogermanisches Etymologisches Wörterbuch*. Vol. 2. Bern: Francke.
- Renfrew, C. (1990). *Archaeology and Language: The Puzzle of Indo-European Origins*. Cambridge: Cambridge University Press.
- Ringe, D. (2006). *From Proto-Indo-European to Proto-Germanic*. Oxford: Oxford University Press.
- Sadori, L., Susanna, F., and Persiani, C. (2006). Archaeobotanical data and crop storage evidence from an early Bronze Age 2 burnt house at Arslantepe, Malatya, Turkey. *Veget. Hist. Archaeobot.* 15, 205–215. doi: 10.1007/s00334-005-0029-3
- Sturtevant, E. H. (1910). Studies in Greek noun-formation: labial termination. *Class. Philol.* 5, 326–356. doi: 10.1086/359423
- Vasmer, M. (1953). *Russisches Etymologisches Wörterbuch*, Vols. I–III. Heidelberg: C. Winter.
- Vavilov, N. I. (1987). *Origin and Geography of Cultivated Plants*, Edited by V. F. Dorofeyev and Translated by D. Löve. Cambridge: Cambridge University Press.
- Zohary, D., Hopf, M., and Weiss, E. (2012). *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford: Oxford University Press.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Dugan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Potential of Legume–Brassica Intercrops for Forage Production and Green Manure: Encouragements from a Temperate Southeast European Environment

Ana M. Jeromela^{1*}, Aleksandar M. Mikić², Svetlana Vujić³, Branko Čupina³, Đorđe Krstić³, Aleksandra Dimitrijević⁴, Sanja Vasiljević², Vojislav Mihailović², Sandra Cvejić¹ and Dragana Miladinović⁴

¹ Oil Crops Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia, ² Forage Crops Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia, ³ Department of Field and Vegetable Crops, Faculty of Agriculture, University of Novi Sad, Novi Sad, Serbia, ⁴ Biotechnology Department, Institute of Field and Vegetable Crops, Novi Sad, Serbia

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico (UNAM), Mexico

Reviewed by:

Vicky M. Temperton,
Lüneburg University, Germany
Ping Wan,
China Agricultural University, China

*Correspondence:

Ana M. Jeromela
ana.jeromela@ifvcns.ns.ac.rs

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 30 November 2016

Accepted: 20 February 2017

Published: 07 March 2017

Citation:

Jeromela AM, Mikić AM, Vujić S, Čupina B, Krstić Đ, Dimitrijević A, Vasiljević S, Mihailović V, Cvejić S and Miladinović D (2017) Potential of Legume–Brassica Intercrops for Forage Production and Green Manure: Encouragements from a Temperate Southeast European Environment. *Front. Plant Sci.* 8:312. doi: 10.3389/fpls.2017.00312

Legumes and brassicas have much in common: importance in agricultural history, rich biodiversity, numerous forms of use, high adaptability to diverse farming designs, and various non-food applications. Rare available resources demonstrate intercropping legumes and brassicas as beneficial to both, especially for the latter, profiting from better nitrogen nutrition. Our team aimed at designing a scheme of the intercrops of autumn- and spring-sown annual legumes with brassicas for ruminant feeding and green manure, and has carried out a set of field trials in a temperate Southeast European environment and during the past decade, aimed at assessing their potential for yields of forage dry matter and aboveground biomass nitrogen and their economic reliability via land equivalent ratio. This review provides a cross-view of the most important deliverables of our applied research, including eight annual legume crops and six brassica species, demonstrating that nearly all the intercrops were economically reliable, as well as that those involving hairy vetch, Hungarian vetch, Narbonne vetch and pea on one side, and fodder kale and rapeseed on the other, were most productive in both manners. Feeling encouraged that this pioneering study may stimulate similar analyses in other environments and that intercropping annual legume and brassicas may play a large-scale role in diverse cropping systems, our team is heading a detailed examination of various extended research.

Keywords: aboveground biomass nitrogen yield, annual legumes, brassicas, forage dry matter yield, intercropping, land equivalent ratio

LEGUMES vs. BRASSICAS OR LEGUMES AND BRASSICAS?

With more than 700 genera and about 20,000 species, legumes (Fabaceae Lindl., syn. Leguminosae Juss.) are considered the third richest plant family (Lewin, 2005). They are renowned for having a remarkable place in global biodiversity, with countless members that are currently wild or semi-domesticated, but with great potential for cultivation. Legumes also provide diverse environment-friendly services, especially in the form of restoring, maintaining or enriching soil fertility through symbiotic nitrogen fixation. Due to their versatile nature, they play a manifold role in various food

and feed production systems as the main, second, cover or cash crop, easily fitting in contrasting and complex agronomic patterns. In numerous regions throughout the world, legumes are staple crops (Vaz Patto et al., 2015), providing local human population and animals with everyday protein-rich diets (Gepts et al., 2005). All these facts endorse legumes with a flattering position of one of the most pivotal components of contemporary and future sustainable agriculture designs (Rubiales and Mikić, 2015).

Although comprising a significantly lesser number of around 370 genera and nearly 4,100 species of mainly herbaceous annuals, biennials and perennials, warm season shrubs and trees (Bremer et al., 2009), Brassicaceae Burnett (syn. Cruciferae Juss.) is a plant family that is by all means not poorer than legumes in economically important crops. Its most remarkable members include rapeseed (*Brassica napus* L.), with its globally widespread cultivar type with low content of erucic acid named *canola* (Lin et al., 2013), cabbage in both broad and narrow sense and with numerous conspecific taxa or variety groups (*Brassica oleracea* L.), and several species of mustards (*Brassica* spp. and *Sinapis* spp.). This family is also a home of thale cress (*Arabidopsis thaliana* L.), with a doubtless historical significance for comparative genetics and genomics and syntenic research, being the first broad-scale studied and the most significant model plant species (Meinke et al., 1998).

Legumes and brassicas have much in common. Both comprise the most ancient domesticated plants in the world, such as chickpea (*Cicer arietinum* L.), lentil (*Lens culinaris* Medik.), pea (*Pisum sativum* L.) or bitter vetch [*Vicia ervilia* (L.) Willd.], *Brassica rapa*, and few *Sinapis* species (Zohary et al., 2012). They also share a characteristic of being not easily preserved as cereals and thus representing demanding archaeobotanical material: high content of protein in legumes and high content of oil in brassicas lead to a much higher degree of degradability in comparison to their fellow starch-rich crops belonging to the family Poaceae Barnhart and an impression that they were drastically less used by the humans in the past. One of the best evidenced findings related to brassicas comes from the famous Neolithic city of Çatalhöyük East, in Asia Minor, where mustards dated at 6,200–6,600 BC were used for ordinary use as oil and spice and for religious ceremonies 8,500 years ago (Fairbairn et al., 2007), along with cereals, grain legumes, and other plants. Apart from these links from deep past, legumes and brassicas today share a vast number of agricultural, economic, and social features. Both are widely present in many local wild floras and offer abundant gene pools for improving the existing and introducing novel cultivated species. Legumes and brassicas are highly adaptable to diverse cropping systems, due to the existence of autumn- and spring-sown cultivars and wide amplitudes of growing season length. They also comprise numerous multi-purpose crops, cultivated for grain, forage, green manure and are used as in human diets or animal feeding, non-food and fuel industry, pharmacy and medicine and ornamental purposes (Mikić et al., 2016).

The presented facts are our answer to the question from the title of this section. It has often been posed as a consequence of very short-term and solely profit-driven motifs, preferring,

for instance, canola to pea and leading to a conclusion that growing one excludes the other or that they always compete for the same field. Rejecting what we deem an artificial antagonism, we declare: “Not legumes *or* brassicas, but legumes *and* brassicas” or, moreover, “legumes *with* brassicas.” This was the basis of the research we have been carrying out throughout the past decade and the results are presented in the forthcoming paragraphs.

TOGETHER IN FEEDING RUMINANTS AND SOIL

A number of studies on intercropping demonstrate its positive effects on crop productivity, weed, pest, and disease control, yield stability, increased resource availability and exploitation, reflecting a renewed interest in this ancient practice. This interest arose from a growing awareness of the environmental problems in modern agriculture associated with excessive use of synthetic fertilisers and pesticides. Due to legumes ability of nitrogen fixation, niche complementarity and positive legumes–cereals interactions, such intercropping systems are known to produce increase in biomass relative to monoculture, a phenomenon described as overyielding (Hector et al., 1999; Beckage and Gross, 2006). Li et al. (2007) showed that faba bean can facilitate mobilisation and uptake of deficient phosphorus by cereals resulting in overyielding. This beneficial interspecific interaction (facilitation) can be based on complementary and more complete sharing and exploitation of other resources, such as solar radiation, water, and soil (Brooker et al., 2015).

In comparison to the accumulated knowledge and available literary resources related to the intercrops of legumes and cereals, it may regretfully be noticed that very little is known on both basic and applied aspects of the mixtures of legumes and brassicas (Mikić et al., 2012). The potential of legume–brassica intercropping has been investigated for various purposes in Europe, Asia, and North America, reflecting the endeavours of researchers to address problems and meet the farmers’ needs specific for these different agro-environments. In France, Sweden, and Canada, the weeds in oilseed rape are tackled by intercropping frost-sensitive legume as living mulch (Bergkvist, 2003; Thériault et al., 2009; Cadoux et al., 2015; Lorin et al., 2015). The advantage of intercropping annual legumes with cabbage and cauliflower in Turkey (Yildirim and Guvenc, 2005, 2006) and broccoli in the USA (Coolman and Hoyt, 1993) over sole crops, were demonstrated in vegetable production. In India, intercropping mustard and annual legumes for oil and food, increases yield stability, nutrient availability, and water use efficiency and provides monetary advantage over sole crops (Banik et al., 2000; Singh et al., 2010; Devi et al., 2014). In several studies, dealing mostly with soil nutrition aspects, it was commonly assessed that a brassica component may profit from being intercropped with a legume. In a series of rhizotron trials, including the intercrops of faba bean (*Vicia faba* L.) with rapeseed and common vetch (*Vicia sativa* L.) with cabbage, it was found out that, if intercropped, the ramification distribution along the taproot was different than in pure stands, which reduced the competition, as well as that the

transfer of nitrogen from legumes to brassicas was significant (Cortés-Mora et al., 2010). Also, the presence of the legumes is demonstrated to be able to help reducing nitrogen fertiliser input (Jamont et al., 2013). A vast and complex field trial in Saskatchewan, Canada, brought forth the thorough claims that intercropping canola with pea improved seed yield, nitrogen uptake and net returns, recommending it as one of the most promising intercrop designs for organic farming systems (Malhi, 2012).

In Southeast Europe, legume–other crop intercropping, especially, legumes–cereals intercropping system has a long tradition for forage production for cattle feed, namely cows. A constant farmers' demand for high yield and good quality forage that will provide high and good quality milk production, prompted research on various combinations of legume–non-legume intercropping patterns, including legume–brassica intercrops. Institute of Field and Vegetable Crops, Novi Sad, Serbia and the Department of Field and Vegetable Crops of the Faculty of Agriculture of the University in Novi Sad, Serbia established a concerted set of field trials, designed jointly by the teams of both institutions and carried out at the Experiment Field at Rimski Šančevi in the vicinity of Novi Sad, on a carbonated chernozem soil and in various consecutive seasons during the past 10 years. The examined species included the following annual legumes and brassicas: common vetch (autumn- and spring-sown), grass pea (*Lathyrus sativus* L.), hairy vetch (*Vicia villosa* Roth), Hungarian vetch (*Vicia pannonica* Crantz), Narbonne vetch (*Vicia narbonensis* L.), pea (autumn- and spring-sown), brown mustard [*Brassica juncea* (L.) Czern.], field mustard [*B. rapa* ssp. *oleifera* (DC.) Metzg.], fodder kale (*B. oleracea* L. var. *viridis* L.), rapeseed (autumn- and spring-sown) and white mustard (*Sinapis alba* L.). These species are commonly used for forage and green manure in the Southeast Europe and were chosen for testing various intercrop combinations.

Our research was founded upon a scheme specifically developed for intercropping autumn- and spring-sown annual legumes and brassicas for both forage production and green manure, in other words, immature aboveground biomass cultivation, that is in bloom and not in full grain maturity (Marjanović-Jeromela et al., 2015c). The scheme is a result of long-term observations of pure stands of various annual legumes and brassicas and consists of three segments (**Figure 1**).

1. A prominently profuse growth of aboveground biomass in forage annual legume cultivars, essential for obtaining high yields of both forage dry matter and aboveground biomass nitrogen, is simultaneously and rather often their main disadvantage: with an insufficient mutual support with tendrils, they are prone to lodging before reaching the stage of full bloom, already lose a significant number of primarily developed leaves and suffer from extreme shade and ideal conditions for disease development in the lower half of the stand (**Figure 1**, top row).

2. On the other hand, brassicas are usually sown at a wider row spacing in order to provide an appropriate space for free and plentiful leaf growth development: however, this is often an opportunity for heavy weed infestations almost immediately after the brassica crop emergence, ending with high proportions of

undesirable plants in total aboveground biomass yield and poorer brassica forage quality (**Figure 1**, middle row).

3. If intercropped with each other, forage annual legumes and brassicas have multiple benefits: brassica serves as an additional and essential support for annual legume, resulting in an improved standing ability, preserved photosynthesis-active leaves and enhanced dry matter production and higher forage protein yield, while legume serves as a powerful weed competitor on behalf of brassica and assists it in accumulating nitrogen in its aboveground biomass (**Figure 1**, bottom row).

This scheme may be compared to those developed for intercropping various annual legumes with each other or pea cultivars with different leaf types, where those, such as white lupin (*Lupinus albus* L.) or faba bean and semi-leafless pea act as supporting crops for forage field pea or vetches (*Vicia* spp.) and normal-leafed pea, respectively (Mikić et al., 2015b).

All the genotypes of both annual legumes and brassicas for intercropping were selected according to the same sowing season, namely autumn and spring, and the concurrent time of cutting, that is, full flower in annual legumes and budding in brassicas, as the optimum moment for both forage production and green manure cultivation. The terms we use, *forage dry matter* and *aboveground biomass* (immature, that is, in bloom and not in full grain maturity) are in fact synonyms: the former denotes that it is cut, collected and used for ruminant feeding, while the latter designs its use as incorporated green manure with potential for increasing soil nitrogen content. These rules are also used in adequately choosing the genotypes for establishing mutual legume mixtures (Ćupina et al., 2011b). In all the examined legume–brassica intercrops, the sowing rates of both components were reduced by half in comparison to their pure stands, as is also present in the designs of brassica–cereal and other crop mixtures (Marjanović-Jeromela et al., 2015b; Mihailović et al., 2015b). A careful choice and adjustment of sowing machine enabled joint sowing of legumes and brassicas, thus avoiding double sowing. Comparing to some other systems, such as relayed intercropping, which require inter-row sowing and temporally separated sowing and harvest, these concurrent mechanised operations are less labour demanding and the economic benefits they could bring may stimulate agriculture machinery manufacturers to develop the machinery adapted for intercropping. Such practice, in case that legume–brassica intercrop schemes are considered for wide production, may serve as another strong encouragement for the farmers to embrace it.

The sole crops of both autumn- and spring-sown annual legumes and brassicas in our set of trials confirmed their high potential for forage production and green manure in temperate regions (Ćupina et al., 2011a; Mihailović et al., 2015a). Legumes had the average values of both forage dry matter yield and aboveground biomass nitrogen in comparison to those of brassicas (**Table 1**; Ćupina et al., 2010; Krstić et al., 2011; Antanasović et al., 2012). In the case of forage dry matter, they ranged from 5.9 t ha⁻¹ in Narbonne vetch to 9.6 t ha⁻¹ in hairy vetch, and between 3.7 t ha⁻¹ in brown mustard and 7.5 t ha⁻¹ in fodder kale. Regarding aboveground biomass nitrogen yield, it varied between 171 kg ha⁻¹ in Narbonne vetch and 327 kg ha⁻¹ in grass pea and from 103 kg ha⁻¹ in brown mustard to



FIGURE 1 | A scheme of intercropping legumes with brassicas for forage production and green manure: *Top row*—often sown in wide rows, *Brassicas* suffer from heavy weed infestations; *Middle row*—forage legumes easily fight the weeds, but are quite prone to lodging, with an outcome in partial or complete withering and loss of lower leaves; *Bottom row*—intercropping legumes and brassicas is beneficial for both, providing legumes with mechanical support and assisting brassicas to bring forth its full potential.

TABLE 1 | Average values of forage dry matter yield (t ha^{-1}), aboveground nitrogen yield (kg ha^{-1}), and their land equivalent ratios (LER_{FDMY} and LER_{AGBY}) in a set of trials at Rimski Šančevi from 2005 to 2014 (Ćupina et al., 2013; Mikić et al., 2013, 2015a; Marjanović-Jeromela et al., 2015a).

Season	Pure stand/intercrop	Total forage dry matter yield	LER_{FDMY}	Total aboveground nitrogen yield	LER_{AGBY}
Autumn	Common vetch	8.5	–	258	–
	Hairy vetch	9.6	–	307	–
	Hungarian vetch	6.5	–	214	–
	Pea	9.2	–	275	–
	Field mustard	6.5	–	188	–
	Fodder kale	7.5	–	199	–
	Rapeseed	6.9	–	189	–
	Common vetch + field mustard	9.9	1.21	282	1.07
	Common vetch + fodder kale	8.4	1.05	238	1.05
	Common vetch + rapeseed	8.3	1.08	214	0.97
	Hairy vetch + field mustard	9.3	1.11	379	1.02
	Hairy vetch + fodder kale	9.3	1.06	303	1.13
	Hairy vetch + rapeseed	9.4	1.09	319	1.18
	Hungarian vetch + field mustard	9.3	1.25	224	1.05
	Hungarian vetch + fodder kale	8.0	1.14	198	0.96
	Hungarian vetch + rapeseed	7.4	1.10	162	0.81
	Pea + field mustard	10.5	1.18	272	1.10
	Pea + fodder kale	8.8	1.06	254	1.07
	Pea + rapeseed	8.9	1.10	242	1.03
	$\text{LSD}_{0.05}$	0.8	0.03	31	0.07
Spring	Common vetch	7.6	–	217	–
	Grass pea	8.8	–	327	–
	Narbonne vetch	5.9	–	171	–
	Pea	8.4	–	240	–
	Brown mustard	3.7	–	103	–
	Rapeseed	6.9	–	171	–
	White mustard	4.2	–	189	–
	Common vetch + brown mustard	6.6	1.11	206	1.14
	Common vetch + rapeseed	8.1	1.12	238	1.23
	Common vetch + white mustard	6.9	1.20	216	1.04
	Grass pea + brown mustard	8.4	1.02	404	1.01
	Grass pea + rapeseed	8.6	1.07	302	1.16
	Grass pea + white mustard	8.0	1.13	352	1.18
	Narbonne vetch + brown mustard	6.6	1.25	198	1.28
	Narbonne vetch + rapeseed	6.5	1.21	199	1.16
	Narbonne vetch + white mustard	6.3	1.23	178	1.00
	Pea + brown mustard	8.0	1.18	249	1.18
	Pea + rapeseed	8.6	1.12	254	1.23
	Pea + white mustard	8.0	1.25	245	1.09
	$\text{LSD}_{0.05}$	0.8	0.09	44	0.06

$\text{LSD}_{0.05}$ = Least significant difference at $p \leq 0.05$.

199 kg ha^{-1} in fodder kale (Ćupina et al., 2013; Mikić et al., 2013). The difference between the trends in forage dry matter yield and aboveground biomass nitrogen yield exist due to the difference in the proportion of crude protein or, in other words, nitrogen in individual species (Ćupina et al., 2014) and is equivalent to those observed in the similar trials with the autumn- and spring-sown intercrops of brassicas and cereals (Mihailović et al., 2014).

The analysis of the average values of forage dry matter yield in the carried out number of field trials show that the autumn-sown intercrops of annual legumes and brassicas were superior to the spring-sown ones, with a variation from 7.4 t ha^{-1} in the

intercrop of Hungarian vetch with rapeseed to 10.5 t ha^{-1} in the intercrop of pea with field mustard and from 6.3 t ha^{-1} in the intercrop of Narbonne vetch and white mustard to 8.6 t ha^{-1} in the intercrops of both grass pea with rapeseed and pea with rapeseed, respectively (Table 1). The intercrops of hairy vetch were most productive in the autumn-sown group, with all three mixtures having forage dry matter yield higher than 9 t ha^{-1} (Ćupina et al., 2013), while, among the spring-sown intercrops, those of grass pea and pea were productive than the other two, both having forage dry matter yield not lower than 8.0 t ha^{-1} (Mikić et al., 2013).

The results related to the average values of aboveground biomass nitrogen yield were generally characterised by a similar trend to that of forage dry matter yield, but more in the autumn-sown group than the spring-sown one: within the former, it was the intercrops of hairy vetch that also had the highest green manure potential, with more than 379 kg ha⁻¹ in all three cases, but, within the latter, it was grass pea that was the most productive, with also more than 300 kg ha⁻¹ (Table 1). This could be, at least partially, explained by a rather great proportion of grass pea in the total aboveground biomass and thus contributing more to the average aboveground biomass nitrogen yield, with the data on these parameters too broad to be shown in this review. The range of average aboveground biomass nitrogen yield among the autumn-sown intercrops was from 162 kg ha⁻¹ in the intercrop of Hungarian vetch with rapeseed to 379 kg ha⁻¹ in the intercrop of hairy vetch with field mustard (Marjanović-Jeromela et al., 2015a), while among the spring-sown intercrops it was from 178 kg ha⁻¹ in the intercrop of Narbonne vetch with rapeseed to 404 kg ha⁻¹ in the intercrop of grass pea with brown mustard (Mikić et al., 2015a).

The land equivalent ratio (LER) is a widely used relative indicator of economic reliability of an intercrop, unlike yield as an absolute one. It is calculated on the basis of the yield of each component in an intercrop and in its pure stand; if surpassing 1.00, an intercrop is considered economically reliable (Mikić et al., 2015b). An overview of the accumulated results related to the LER values of both forage dry matter yield (LER_{FDMY}) and aboveground biomass nitrogen yield (LER_{AGBY}) shows that all the intercrops had higher LER_{FDMY} than 1.00, as well as that a very few intercrops had LER_{AGBY} values lower than 1.00 (Table 1). The intercrops of Hungarian vetch with field mustard, Narbonne vetch with brown mustard and pea with white mustard had the highest average values of LER_{FDMY} (Ćupina et al., 2013), 1.25 in all three, while the intercrop of Narbonne vetch with brown mustard had the highest average value of LER_{AGBY}, 1.28. Among the autumn-sown intercrops, the highest values of LER_{FDMY} and LER_{AGBY} were in those of Hungarian vetch and hairy vetch, respectively (Mikić et al., 2015a), while among the spring-sown intercrops, the highest values of LER_{FDMY} and LER_{AGBY} were in those of Narbonne vetch (Mikić et al., 2013) and pea (Marjanović-Jeromela et al., 2015a).

REFERENCES

- Antanasović, S., Ćupina, B., Krstić, Đ., Manojlović, M., Čabilovski, R., Marjanović-Jeromela, A., et al. (2012). Potential of autumn-sown rapeseed (*Brassica napus*) as a green manure crop. *Crucif. Newsl.* 31, 26–28.
- Banik, P., Sasmal, T., Ghosal, P. K., and Bagchi, D. K. (2000). Evaluation of mustard (*Brassica campestris* var. Toria) and legume intercropping under 1:1 and 2:1 row replacement series systems. *J. Agron. Crop Sci.* 185, 9–14. doi: 10.1046/j.1439-037X.2000.00388.x
- Beckage, B., and Gross, L. J. (2006). Overyielding and species diversity: what should we expect? *New Phytol.* 172, 140–148.
- Bergkvist, G. (2003). Influence of white clover traits on biomass and yield in winter wheat or winter oilseed rape-clover intercrops. *Biol. Agric. Hortic.* 21, 151–164. doi: 10.1080/01448765.2003.9755259
- Bremer, B., Bremer, K., Chase, M., Fay, M., Reveal, J., Soltis, D., et al. (2009). An update of the Angiosperm Phylogeny Group classification for the orders

Legume + Brassica Perspectives

The authors are fully aware of a multitude of issues stemming out from the achieved, accumulated and presented results and all the arguments possibly raised by a reader. However, all of us being practical agronomists and breeders, we aimed first at firmly demonstrating that our schemes for intercropping annual legume and brassicas answer the most important demand by the farmers: that the forage yield is higher than in pure stands. Having a positive answer, along with the good results in relation to the economic reliability, we feel encouraged to proceed with a detailed examination of various extended research, such as assessing the intercrop chemical composition, examining the relationship towards various forms of abiotic and biotic stress and, especially, studying the complex underground aspects of nutrition and allelopathy. We remain convinced that this pioneering review may stimulate similar analyses in other environments, as well as that intercropping annual legume and brassicas may secure its place in diverse farming systems and on a larger scale based on its many advantages, not least the use of local legumes and *Brassica* species, thus strengthening local economies as well as promoting conservation, since no fertilisers are required.

AUTHOR CONTRIBUTIONS

AJ, AM, BĆ, and VM planned and designed the experiments; AJ, AM, ĐK, AD, and SC: Performed the research; AJ, AM, BĆ, VM, SV, and DM contributed to interpretation and analysis of results; AJ, AM, SV, BĆ, ĐK, AD, SVa, VM, SC, and DM wrote and approved the manuscript.

FUNDING

This work was funded by projects TR31016, TR31024, and TR31025, co-funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia and project 114-451-2180/2016-01, co-funded by Provincial Secretariat for Higher Education and Scientific Research, Autonomous Province of Vojvodina, Republic of Serbia.

and families of flowering plants: APG III. *Bot. J. Linn. Soc.* 161, 105–121. doi: 10.1111/j.1095-8339.2009.00996.x

- Brooker, R. W., Bennett, A. E., Cong, W. F., Daniell, T. J., George, T. S., Hallett, P. D., et al. (2015). Improving intercropping: a synthesis of research in agronomy, plant physiology and ecology. *New Phytol.* 206, 107–117. doi: 10.1111/nph.13132
- Cadoux, S., Sauzet, G., Valantin-Morison, M., Pontet, C., Champolivier, L., Céline, R., et al. (2015). Intercropping frost-sensitive legume crops with winter oilseed rape reduces weed competition, insect damage, and improves nitrogen use efficiency. *OCL* 22, D302. doi: 10.1051/ocl/2015014
- Coolman, R. M., and Hoyt, G. D. (1993). Increasing sustainability by intercropping. *HortTechnology* 3, 309–312.
- Cortés-Mora, A. F., Piva, G., Jamont, M., and Fustec, J. (2010). Niche separation and nitrogen transfer in *Brassica*-legume intercrops. *Ratar. Povrt.* 47, 581–586.
- Ćupina, B., Krstić, Đ., Antanasović, S., Erić, P., Manojlović, M., Čabilovski, R., et al. (2010). Potential of fodder kale (*Brassica*

- oleracea L. var. *viridis* L.) as a green manure crop. *Crucif. Newsl.* 29, 10–11.
- Čupina, B., Manojlović, M., Krstić, Đ., Čabilovski, R., Mikić, A., Ignjatović-Čupina, A., et al. (2011a). Effect of winter cover crops on the dynamics of soil mineral nitrogen and yield and quality of Sudan grass (*Sorghum bicolor* (L.) Moench). *Austr. J. Crop Sci.* 5, 839–845.
- Čupina, B., Mikić, A., Marjanović-Jeromela, A., Krstić, Đ., Antanasović, S., Erić, P., et al. (2013). Intercropping autumn-grown brassicas with legumes for forage production. *Crucif. Newsl.* 32, 17–19.
- Čupina, B., Mikić, A., Marjanović-Jeromela, A., Živanov, D., Krstić, Đ., Antanasović, S., et al. (2014). Forage dry matter crude protein yield in the intercrops of spring-sown brassicas with legumes. *Crucif. Newsl.* 33, 13–15.
- Čupina, B., Mikić, A., Stoddard, F. L., Krstić, Đ., Justes, E., Bedoussac, L., et al. (2011b). “Mutual legume intercropping for forage production in temperate regions,” in *Sustainable Agriculture Reviews 7: Genetics, Biofuels and Local Farming Systems*, ed. E. Lichtfouse (Dordrecht: Springer), 347–365.
- Devi, K. N., Shamurailatpam, D., Singh, T. B., Athokpam, H. S., Singh, N. B., Singh, N. G., et al. (2014). Performance of lentil (*Lens culinaris* M.) and mustard (*Brassica juncea* L.) intercropping under rainfed conditions. *Aust. J. Crop Sci.* 8, 284–289.
- Fairbairn, A., Martinoli, D., Butler, A., and Hillman, G. (2007). Wild plant seed storage at Neolithic Çatalhöyük east. *Turkey. Veg. Hist. Archaeobot.* 16, 467–479. doi: 10.1007/s00334-006-0069-3
- Gepts, P., Beavis, W. D., Brummer, E. C., Shoemaker, R. C., Stalker, H. T., Weeden, N. F., et al. (2005). Legumes as a model plant family. Genomics for food and feed report of the cross-legume advances through genomics conference. *Plant Physiol.* 137, 1228–1235. doi: 10.1104/pp.105.060871
- Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M. C., Diemer, M., Dimitrakopoulos, P. G., et al. (1999). Plant diversity and productivity experiments in European grasslands. *Science* 286, 1123–1127. doi: 10.1126/science.286.5442.1123
- Jamont, M., Piva, G., and Fustec, J. (2013). Sharing N resources in the early growth of rapeseed intercropped with faba bean: does N transfer matter? *Plant Soil* 371, 641–653. doi: 10.1007/s11104-013-1712-2
- Krstić, Đ., Mikić, A., Čupina, B., Antanasović, S., Mihailović, V., Erić, P., et al. (2011). Forage yields in the intercrops of pea with other cool season annual legumes. *Pisum Genet.* 43, 29–32.
- Lewis, G. P. (2005). *Legumes of the World*. Kew: Royal Botanic Gardens.
- Li, L., Li, S. M., Sun, J. H., Zhou, L. L., Bao, X. G., Zhang, H. G., et al. (2007). Diversity enhances agricultural productivity via rhizosphere phosphorus facilitation on phosphorus-deficient soils. *Proc. Natl. Acad. Sci. U.S.A.* 104, 11192–11196. doi: 10.1073/pnas.0704591104
- Lin, L., Allemekinders, H., Dansby, A., Campbell, L., Durance-Tod, S., Berger, A., et al. (2013). Evidence of health benefits of canola oil. *Nutr. Rev.* 71, 370–385. doi: 10.1111/nure.12033
- Lorin, M., Jeuffroy, M. H., Butier, A., and Valantin-Morison, M. (2015). Undersowing winter oilseed rape with frost-sensitive legume living mulches to improve weed control. *Eur. J. Agron.* 71, 96–105. doi: 10.1016/j.eja.2015.09.001
- Malhi, S. S. (2012). Improving crop yield, N uptake and economic returns by intercropping barley or canola with pea. *Agric. Sci.* 3, 1023–1033. doi: 10.4236/as.2012.38124
- Marjanović-Jeromela, A., Mikić, A., Čupina, B., Antanasović, S., Erić, P., Terzić, S., et al. (2015a). Intercropping spring-sown brassicas with annual legumes for green manure. *Crucif. Newsl.* 34, 8–10.
- Marjanović-Jeromela, A., Mikić, A., Mihailović, V., Aćin, V., Kondić-Špika, A., Hristov, N., et al. (2015b). Forage dry matter crude protein yield in the intercrops of spring-sown brassicas with cereals. *Crucif. Newsl.* 34, 17–19.
- Marjanović-Jeromela, A., Mikić, A., Terzić, S., Vujić, S., Mihailović, V., Čupina, B., et al. (2015c). Intercropping annual legumes and brassicas for forage production. *Legume Perspect.* 8, 31–32.
- Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., and Koornneef, M. (1998). *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282, 662–682. doi: 10.1126/science.282.5389.662
- Mihailović, V., Mikić, A., Čupina, B., Marjanović-Jeromela, A., Marinković, R., Živanov, D., et al. (2014). Intercropping spring-sown brassicas with cereals for forage production. *Crucif. Newsl.* 33, 16–18.
- Mihailović, V., Mikić, A., Erić, P., Čupina, B., Vasiljević, S., Karagić, Đ., et al. (2015a). Annual forage legumes in temperate South-East Europe. *Legume Perspect.* 12, 39–40.
- Mihailović, V., Mikić, A., Marjanović-Jeromela, A., Miroslavljević, M., Hristov, N., Kondić-Špika, A., et al. (2015b). Forage dry matter crude protein yield in the intercrops of autumn-sown brassicas with cereals. *Crucif. Newsl.* 34, 20–22.
- Mikić, A., Antanasović, S., Čupina, B., Marjanović-Jeromela, A., Erić, P., Jakšić, S., et al. (2015a). Intercropping autumn-sown brassicas with annual legumes for green manure. *Crucif. Newsl.* 34, 11–13.
- Mikić, A., Čupina, B., Mihailović, V., Krstić, Đ., Đorđević, V., Perić, V., et al. (2012). “Forage legume intercropping in temperate regions: models and ideotypes,” in *Sustainable Agriculture Reviews 11*, ed. E. Lichtfouse (Dordrecht: Springer), 161–182.
- Mikić, A., Čupina, B., Rubiales, D., Mihailović, V., Šarūnaitė, L., Fustec, J., et al. (2015b). Developments, and perspectives of mutual legume intercropping. *Adv. Agron.* 130, 337–419. doi: 10.1016/bs.agron.2014.10.004
- Mikić, A., Čupina, B., Mihailović, V., Marjanović-Jeromela, A., Antanasović, S., Krstić, Đ., et al. (2013). Intercropping spring-grown brassicas and legumes for forage production. *Crucif. Newsl.* 32, 14–16.
- Mikić, A., Marjanović Jeromela, A., and Terzić, S. (2016). Brassicaceae are simply beautiful! *Crucif. Newsl.* 35, 36–39.
- Rubiales, R., and Mikić, A. (2015). Introduction: legumes in sustainable agriculture. *Crit. Rev. Plant Sci.* 34, 2–3. doi: 10.1080/07352689.2014.897896
- Singh, R. K., Kumar, H., and Singh, A. K. (2010). Brassica based intercropping systems - A review. *Agric. Rev.* 31, 253–266.
- Thériault, F., Stewart, K. A., and Seguin, P. (2009). Use of perennial legumes living mulches and green manures for the fertilization of organic broccoli. *Int. J. Veg. Sci.* 15, 142–157. doi: 10.1080/19315260802598896
- Vaz Patto, M. C., Amarowicz, R., Aryee, A. N., Boye, J. I., Chung, H. J., Martín-Cabrejas, M. A., et al. (2015). Achievements and challenges in improving the nutritional quality of food legumes. *Crit. Rev. Plant Sci.* 34, 105–143. doi: 10.1080/07352689.2014.897907
- Yildirim, I., and Guvenc, E. (2005). Intercropping based on cauliflower: more productive, profitable and highly sustainable. *Eur. J. Agron.* 22, 11–18. doi: 10.1016/j.eja.2003.11.003
- Yildirim, I., and Guvenc, E. (2006). Increasing productivity with intercropping systems in cabbage production. *J. Sustain. Agric.* 28, 29–44. doi: 10.1300/J064v28n04_04
- Zohary, D., Hopf, M., and Weiss, E. (2012). *Domestication of Plants in the Old World: The Origin and Spread of Domesticated Plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford: Oxford University Press. doi: 10.1093/acprof:osobl/9780199549061.001.0001

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Jeromela, Mikić, Vujić, Čupina, Krstić, Dimitrijević, Vasiljević, Mihailović, Cvejić and Miladinović. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Influence of Harvest Aid Herbicides on Seed Germination, Seedling Vigor and Milling Quality Traits of Red Lentil (*Lens culinaris* L.)

Maya Subedi*, Christian J. Willenborg and Albert Vandenberg*

Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica - Universidade Nova de
Lisboa, Portugal

Reviewed by:

Alma Balestrazzi,
University of Pavia, Italy
Robert John French,
Department of Agriculture and Food,
Australia

*Correspondence:

Albert Vandenberg
bert.vandenberg@usask.ca
Maya Subedi
mas248@mail.usask.ca

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 06 December 2016

Accepted: 20 February 2017

Published: 14 March 2017

Citation:

Subedi M, Willenborg CJ and
Vandenberg A (2017) Influence of
Harvest Aid Herbicides on Seed
Germination, Seedling Vigor and
Milling Quality Traits of Red Lentil (*Lens
culinaris* L.). *Front. Plant Sci.* 8:311.
doi: 10.3389/fpls.2017.00311

Most red lentil produced worldwide is consumed in dehulled form, and post-harvest milling and splitting qualities are major concerns in the secondary processing industry. Lentil producers in northern temperate regions usually apply pre-harvest desiccants as harvest aids to accelerate the lentil crop drying process and facilitate harvesting operations. This paper reports on field studies conducted at Scott and Saskatoon, Saskatchewan, Canada in the 2012 and 2013 cropping seasons to evaluate whether herbicides applied as harvest aids alone or tank mixed with glyphosate affect seed germination, seedling vigor, milling, and splitting qualities. The site-year by desiccant treatment interaction for seed germination, vigor, and milling recovery yields was significant. Glyphosate applied alone or as tank mix with other herbicides (except diquat) reduced seed germination and seedling vigor at Saskatoon and Scott in 2012 only. Pyraflufen-ethyl (20 g ai ha⁻¹) applied with glyphosate as well as saflufenacil (36 g ai ha⁻¹) decreased dehulling efficiency, while saflufenacil and/or glufosinate with glyphosate reduced milling recovery and footfall recovery, although these effects were inconsistent. Application of diquat alone or in combination with glyphosate exhibited more consistent dehulling efficiency gains and increases in milling recovery yield. Significant but negative associations were observed between glyphosate residue in seeds and seed germination ($r = -0.84, p < 0.001$), seed vigor ($r = -0.62, p < 0.001$), dehulling efficiency ($r = -0.55, p < 0.001$), and milling recovery ($r = -0.62, p < 0.001$). These results indicate application of diquat alone or in combination with glyphosate may be a preferred option for lentil growers to improve milling recovery yield.

Keywords: dehulling efficiency, footfall recovery, milling recovery, pre-harvest aid, desiccant, lentil

INTRODUCTION

Lentil (*Lens culinaris* Medik.) is a valuable grain legume crop that is a good source of dietary protein, fiber, complex carbohydrates, and minerals (Jood et al., 1998; Xu and Chang, 2009; DellaValle et al., 2013). The Western Canadian Prairies is the world's major lentil producing and exporting region, with a current production area of 2.3 million ha (Statistics Canada, 2016). The main destinations for red lentil are India, Turkey, the United Arab Emirates, and the European Union (Statistics Canada, 2016). About 90% of red lentils are consumed in dehulled form after removal of the seed coat through an abrasive dehulling process that improves the taste of cooked lentil by removing

anti-nutritional factors, such as polyphenols and tannins, which are mostly retained in the seed coat fraction (Singh and Singh, 1992; Wang, 2008; DellaValle et al., 2013). Efficient loosening of the seed coat during dehulling process of red lentil is vital for the milling industry. The market value of red lentil largely depends on milling quality, which in turn depends on genetics but also on agronomic practices, such as the application of pre-harvest desiccants and growing environment (Ramakrishnaiah and Kurien, 1983; Wang, 2008).

Lentils harvested under low humidity and high temperature conditions in Australia, Mediterranean and sub-tropical savannah climates are more efficiently dehulled than those harvested in a temperate climate (Brand, 2008). Northern temperate prairie environments experience much different climatic conditions compared to global competitors who tend to harvest during hot dry conditions. Indeed, red lentils produced in northern temperate prairie regions generally have a higher moisture content at harvest and different physical characteristics than those grown in Mediterranean and sub-tropical savannahs. The climate, genetic base, and agronomic practices for red lentil grown in northern temperate climates are unique, and may result in poor milling recovery compared to lentil produced and milled in dry environments (Vandenberg, 2009).

In Western Canada, the environmental variation within fields and the indeterminate nature of crop growth means that lentil growers usually apply pre-harvest desiccants to optimize harvest conditions on the variable landscape. Lentils are considered sufficiently mature for desiccation with harvesting aids when 80% of the pods in the lower third of the canopy have turned from green to yellow or brown (Saskatchewan Pulse Growers, 2015). Desiccant chemistry and application timing are crucial as they may cause loss of yield and quality (Bennett and Shaw, 2000a; Wilson and Smith, 2002; Zhang et al., 2016). In Western Canada, the herbicides registered as harvest aids for lentil and other pulses include diquat, glyphosate, saflufenacil, glufosinate, and flumioxazin (Risula, 2014). These desiccants have different modes of action and chemistry and, therefore, may differentially affect post-harvest seed quality.

Diquat is a quick acting contact herbicide that is traditionally used as a harvest aid for lentil. It rapidly and quickly dries plant tissues within few days of application and has no or low translocation in plants (Cobb and Reade, 2010; Zhang et al., 2016). However, pre-harvest application of diquat affects milling recovery yield and other post-harvest seed qualities in lentil when it is applied too early (Bruce, 2008). Glyphosate, a popular herbicide product, is used in lentil production to control late emerging annual weeds and acts as a desiccant. It is also used as a desiccant in common bean production in Canada (McNaughton et al., 2015). Seeds are a major photosynthate sink during maturation (Cakmak et al., 2009); glyphosate is translocated mainly through phloem and distributed throughout the plant, and seed germination and vigor may be reduced if seeds accumulate too much glyphosate residue (Clay and Griffin, 2000; Zhang et al., 2016). Glufosinate and saflufenacil are newly registered as desiccants for lentil crops (Risula, 2014). Saflufenacil is a weak acid herbicide used for broadleaf weed control in soybean, corn, and other crops. It moves both

acropetally and basipetally in plants and affects symptoms similar to other contact herbicides (Soltani et al., 2010). Glufosinate can translocate within plants, but has rapid phytotoxicity that limits its mobility (Grossmann et al., 2010; Soltani et al., 2013). Pyraflufen-ethyl and flumioxazin are also potential harvest aids for lentil crops in Western Canada (Risula, 2014; Zhang et al., 2016) and are commonly used as desiccants in cotton (Griffin et al., 2010) and common bean (Soltani et al., 2013). Overall, minimal research has been conducted to assess the impact of the complete spectrum of harvest aids (e.g., diquat, pyraflufen, glufosinate ammonium, flumioxazin, and saflufenacil) alone or in combination with glyphosate with respect to their effects on seed biology and post-harvest processing of lentil. This study was conducted to evaluate the effect of contact herbicides applied alone or as tank mixtures with glyphosate as harvest aids on milling recovery yield, seed germination, seedling vigor, and other seed quality attributes of red lentil.

MATERIALS AND METHODS

Field Experiments

Harvested seed samples of red lentil (cv. CDC Maxim) were obtained from field experiments conducted by Zhang et al. (2016) in two Saskatchewan locations in 2012 and 2013. These locations (Saskatoon, 52°36' N, 108°84' W, altitude 659.6 m; Scott, 52°09' N, 106°33' W, altitude 505 m) are located in the Dark Brown zone of Chernozemic soils (clay to sandy loam at Saskatoon and silt loam at Scott). The soil organic matter content and pH ranged from 2.4 to 4.5% and 7.5 to 7.9 at Saskatoon and 2.4 to 2.6% and 5.3 to 6.8 at Scott, respectively.

The experimental treatments (Table 1) were arranged in a randomized complete block design (RCBD) with four replicates. Each block consisted of 21 desiccants and an unsprayed control. The desiccants used in this study were pyraflufen-ethyl (10 and 20 g ai ha⁻¹), glufosinate (300 and 600 g ai ha⁻¹), flumioxazin (105 and 210 g ai ha⁻¹), saflufenacil (36 and 50 g ai ha⁻¹) and diquat (208 and 415 g ai ha⁻¹), with each desiccant applied alone or in combination with glyphosate (900 g ae ha⁻¹). The list of treatments is presented in Table 1. All desiccant treatments were applied with the recommended adjuvant, either Merge® (50% surfactant; 50% petroleum hydrocarbons solvent) or Agral 90® (90% nonylphenoxy polyethoxy ethanol) with an air pressurized tractor mounted sprayer equipped with shielding (110–015 AirMix nozzles, 275 kpa, 45 cm spacing at Saskatoon and with a CO₂ pressurized bicycle sprayer (110–003 AirMix nozzles, 276 kpa, 25 cm) at Scott. Both sprayers were calibrated to deliver 200 L ha⁻¹ of spray solution. Prior to crop harvest aid applications, the seed moisture content was determined by picking and bulking a few seeds from two border plots to create a composite sample, which was then dried at 90°C for 24 h. Treatments were applied to the lentil crop at approximately 30% seed moisture content at the recommended stage (when lower seeds rattled and pods started turning brown, and middle pods were yellow to brown). The crops were harvested 21 days after desiccant application.

The red cotyledon lentil cultivar CDC Maxim (40–45 mg seed weight) was chosen for the study because it is the most widely grown cultivar in Canada. It is tolerant to imidazolinone

TABLE 1 | Desiccant (pre-harvest aids) treatments and their rate of application.

Desiccant /Herbicide	Rate (g a.i./a.e. ha ⁻¹)	Surfactants
Untreated check		-
Glyphosate	900	-
Pyraflufen	10	Merge (1% v/v)
	20	Merge (1% v/v)
Pyraflufen+Glyphosate	10 + 900	Merge (1% v/v)
	20 + 900	Merge (1% v/v)
Glufosinate	300	-
	600	-
Glufosinate + Glyphosate	300 + 900	-
	600 + 900	-
Flumioxazin	105	Agral 90 (0.25% v/v)
	210	Agral 90 (0.25% v/v)
Flumioxazin+Glyphosate	105 + 900	Agral 90 (0.25% v/v)
	210 + 900	Agral 90 (0.25% v/v)
Saflufenacil	36	Merge (1% v/v)
	50	Merge (1% v/v)
Saflufenacil + Glyphosate	36 + 900	Merge (0.5% v/v)
	50 + 900	Merge (0.5% v/v)
Diquat	208	Agral (0.1%v/v)
	415	Agral (0.1%v/v)
Diquat+Glyphosate	208 + 900	Agral (0.1%v/v)
	415 + 900	Agral (0.1%v/v)

herbicides. Details of field management are documented by Zhang et al. (2016). Prior to seeding, lentil seeds were inoculated with liquid Nodulator[®] inoculant (*Rhizobium leguminosarum biovar viceae*) at a rate of 2.76 mL kg⁻¹ in 2012, and with Tag Team[®] Granular (*Rhizobium leguminosarum* and *penicillium bilaii*) at a rate of 2.8 kg ha⁻¹ in 2013. Seeds were treated with Apron Maxx RTA (0.73% fludioxonil; 1.10% metalaxyl-M and S-isomer) at a rate of 325 ml per 100 kg of seed before sowing in each site. After treatment, seeds were sown at 3 cm depth with seeding density of 130 seeds m⁻² using a small plot drill equipped with hoe openers spaced at 22 cm between rows. Individual plots were 2.25 m wide and 6 m long at Saskatoon and 2 m wide and 5 m long at Scott, and consisted of six rows at both sites. Sowing was performed in mid-May in each year at each location. A tank mixture of imazamox and imazethapyr (30 g a.i. ha⁻¹) was sprayed between 5 and 6th lentil node stage at Saskatoon and quizalofop-p-ethyl (420 g a.i. ha⁻¹) was sprayed at 4th node stage of lentil for post emergence weed control. Hand weeding was carried out to maintain plots weed free. The fungicides prothioconazole (166 g a.i. ha⁻¹) and boscalid (294 g a.i. ha⁻¹) were applied at Saskatoon and Scott, respectively, to control foliar diseases.

Post-harvest Seed Measurements

Randomly selected 250-seed samples from individual plots were collected, counted using an electronic seed counter (ESC-1 Agriculex Inc., Guelph, ON, Canada), and weighed to determine 1000-seed weight. Seed diameter and thickness were measured using round-hole and slotted-hole sieves (Hossain et al., 2010;

Fedoruk et al., 2013). Seed diameter was measured by passing seed samples of approximately 100 g through a set of 10 round-hole sieves ranging from 5.8 mm (15/64") down to 3.6 mm (9/64") in 0.25 mm (1/64") increments. Seed thickness was measured by passing the same sample through a set of six slotted-hole sieves from 3.1 mm (8/64") down to 2.0 mm (5/64") in 0.2 mm (0.5/64") increments. All samples were shaken for 1 min on a flatbed shaker (Lab Line Instruments, Melrose Park, Illinois, USA). The seed fractions remaining in each round and slotted-hole sieve were weighed. Seed diameter and thickness for each sample were calculated using the following formulas:

$$\text{Percentage on sieve} = \frac{\text{weight of seed in each sieve (g)}}{\text{weight of total sample seed (g)}} \times 100,$$

$$\text{Mean seed diameter} = \left[\frac{\% \text{ of seed weight on round sieve}}{100} \times \text{sieve hole size (mm)} \right],$$

$$\text{Mean seed thickness} = \left[\frac{\% \text{ of seed weight on slotted sieve}}{100} \times \text{sieve hole size (mm)} \right],$$

$$\text{Seed plumpness} = \frac{\text{mean seed thickness}}{\text{mean seed diameter}}.$$

Dehulling Procedure and Separation of Dehulled Fractions

Prior to dehulling, initial moisture content of the seeds was determined using an oven dry method (AACC, 2000). For each whole seed sample, 16 g was dried at 130°C for 20 h, and the weight difference of each sample expressed as moisture percentage.

Lentil seed samples (30 g) that remained in the round (4.47–5.16 mm) and slotted (2.38–2.58 mm) sieves were tempered to 12.5% moisture (Wang, 2005) and then dehulled using a grain testing mill (TM05, Satake Engineering Co., Hiroshima, Japan) fitted with a 36-mesh abrasive wheel rotating at 1100 rpm for 38 s, as described by Wang (2005). After dehulling, milled samples were collected in a paper envelope and then the entire milled sample was weighed and separated into different fractions. For separation, the first sample was screened on Canada standard No. 14 (1.40 mm) and No. 35 mesh (850 μm) sieves. The powder was collected and weighed. The leftover fraction in the No. 14 sieve was passed through an aspiration unit to separate the hull portions. The sample seeds remaining in the aspiration column were further sieved to separate fractions. The remaining lentil seeds without powder and hulls were passed through a No. 6 (6/64") slotted sieve and No. 9 (9/64") round sieve. Whole lentils remaining over the slotted sieve were considered football, lentils remaining in the round sieve were considered split, and material in the pan was considered broken seed. Any whole and split lentils with adhering hulls were separated manually into their respective adhering hulled or dehulled classes. All fractions were weighed and then expressed as a proportion of the total original

milled sample weight. Dehulling efficiency was defined as the percent of un-dehulled whole and split (%) seed relative to total initial sample weight (Wang, 2005; Bruce, 2008); and milling recovery indicated as the percent of dehulled splits and football fractions to total initial sample weight. Football recovery, milling recovery, and dehulling efficiency were calculated according to the following:

$$\begin{aligned}\text{Dehulling efficiency (DE, \%)} &= \left[1 - \frac{(\text{weight of unde-hulled whole seed (g)} + \text{unde-hulled split seed (g)})}{\text{weight of sample seed (g)}} \right] \times 100, \\ \text{Milling recovery (MR, \%)} &= \frac{\text{weight of milled seeds (g)}}{\text{weight of sample seed (g)}} \times 100 \\ \text{Football recovery (FR, \%)} &= \frac{\text{weight of football (un - split intact seed) (g)}}{\text{weight of sample seed (g)}} \times 100.\end{aligned}$$

Germination and Vigor Tests

Seed germination tests were performed at Discovery Seed Labs Ltd. (Saskatoon, SK) using the rolled paper towel method and procedures recommended for lentil seed by the Canadian Food Inspection Agency (CFIA, 2012). Two hundred seeds from each plot in each replication were evenly spaced on two sheets of germination paper and then covered with a moistened paper. Four replication were used. The sheets of paper were rolled and placed in an upright position. The rolled paper sheets were moistened daily by adding water. The temperature was maintained at 20°C. After 7 d, the number of normal seedling and abnormal seedlings, such as number of un—germinate fresh, dormant, hard and dead seed were counted. Then percentage of normal seedling were used to expresses germination percentage.

Seedling vigor was also determined by the standard method developed by Canadian Food Inspection Agency at Discovery Seed Labs Ltd., using 200-seed samples at 5°C for 7 d. The number of normal and vigorous seedlings were counted 7 d after emergence and expressed as a percentage.

Glyphosate Residue Content

The glyphosate residue data, reported by Zhang et al. (2016) using high performance liquid chromatography (HPLC) column switching and post-column derivatization with fluorescence detection to determine glyphosate (at ALS Laboratories, Edmonton, AB, Canada), were used for correlation analysis among selected traits. Glyphosate residue content was analyzed using both treated and un-treated seeds. Each 250 g seed sample was collected at 7 DAA from border rows, cleaned, placed into plastic bags and kept in a freezer at −20°C until all samples were collected. Samples were sent to ALS laboratory in Edmonton, AB, Canada. Using a standardized process provided by ALS Laboratories, high performance liquid chromatography (HPLC) using column switching and post-column derivatization with fluorescence detection was employed to determine glyphosate and AMPA residue. Briefly, a mixture of 150 ml of 0.1 M hydrochloric acid and 50 ml of dichloromethane was added to ground samples. The solution was homogenized for 1 min with a polytron, and centrifuged at 5000 RPM for 10 min. The aqueous

layer of this solution (100 ml) was decanted to a flask and diluted with deionized water to 350 ml, and eluted through a Chelex 100 resin column at 2 drops per second. The wall of this column was then washed with 50 ml of deionized water and 100 ml of 0.2 M hydrochloric acid. All the eluent was discarded. Following this, 7 ml of 6 M hydrochloric acid was added to the column, and the eluent was discarded. 25 ml of 6 M hydrochloric acid was added

again to the column, and with the eluent collected, mixed with 11 ml of concentrated hydrochloric acid and applied to a AG1-X8 resin column to remove excess iron. After the eluent entered the AG1-X8 resin column, the column was rinsed with 10 ml of 6 M hydrochloric acid, and the eluent was concentrated on a rotary evaporator. The extract of glyphosate and AMPA was then determined with an HPLC equipped with a fluorescence detector. Differential retention time was used to distinguish between glyphosate and AMPA, with a limit of detection of 0.020 ppm for both compounds.

Data Analyses

Normality and homogeneity tests were performed using residual data through the PROC UNIVARIATE procedure and Levene's test, respectively, prior to using a mixed model in SAS 9.3 (SAS Institute, Inc., Cary, NC, USA; SAS, 2015). Data were pooled and analyzed using the PROC MIXED procedure in a Randomized complete block design. The pre-harvest aid (desiccant) treatment was considered a fixed factor, whereas environment (year × location site year), environment × treatment interaction, and blocks were considered random factors. In the mixed model analysis, the significance of the fixed effect was tested using *F*-tests, whereas random effects were tested using a *Z*-test of the variance estimate. The REPEATED/GROUP statement was used to model heterogeneous variance for germination data from 2012 samples because these data did not meet the assumption to use ANOVA even after transformation. The covariance parameter estimation (COVTEST option of PROC MIXED) was used to determine whether or not data might be combined across site-years for analysis. The data were analyzed for each year and location separately for those variables that had significant interactions of site-year with desiccants. Fisher's least significant difference (LSD) was performed for mean separation with a 5% significance level. Additionally, all letter groupings for significance differences were established using PDMIX 800 in SAS (Saxton, 1998). Simple linear contrast estimate was used to compare differences in mean of groups. PROC CORR command in SAS 9.3 was used to analyze correlation among selected traits.

RESULTS

Seed Physical Characteristics: Diameter, Thickness, and Plumpness

Desiccation treatments had a marginal effect on 1000-seed weight and seed moisture content prior to conditioning; therefore, data for these parameters are not presented. And weather information related mean monthly temperature and precipitation during the growing season at each location in 2012 and 2013 are presented in **Table 2**. Irrespective of year or location, glyphosate applied alone or as a tank mix with other herbicides had no significant effect on seed diameter, thickness, or seed plumpness (**Table 3**). Effects on seed physical dimensions were consistent across all site-years, as no significant interactions were observed between desiccant and site-year (**Table 3**). However, contrast analysis showed that the addition of contact herbicides to glyphosate increased seed diameter compared to glyphosate applied alone (**Table 4**). Conversely, application of higher rates of contact herbicides significantly decreased seed diameter by 2%. In contrast, neither addition of contact herbicide with glyphosate nor glyphosate applied alone affected seed thickness or plumpness (**Table 4**). These results suggest that none of the contact herbicides considered, applied alone or in tank mixes with glyphosate, had any adverse effect on seed physical qualities.

Seed Biological Characteristics: Germination and Seedling Vigor

The impact of desiccant treatment on seed germination and seedling vigor varied with growing environment (**Table 3**) so these data were analyzed separately by site year. At Saskatoon in 2012, diquat, pyraflufen, and glufosinate (300 g a.i. ha⁻¹), and flumioxazin did not affect lentil seed germination compared to

the untreated control; in contrast, plots sprayed with glyphosate (900 g a.i. ha⁻¹) alone or with these desiccants as tank-mix with glyphosate except (diquat + glyphosate) resulted in a significant reduction of seed germination compared to the untreated control (**Table 6**). Furthermore, adding other contact herbicides to glyphosate significantly increased seed germination (10.9%) over glyphosate applied alone as did application of lower rates of contact herbicide alone. Similar results were observed at Scott in 2012, where seeds from plots sprayed with glyphosate alone and in combination with pyraflufen (10 or 20 g a.i. ha⁻¹), glufosinate (300 g a.i. ha⁻¹), flumioxazin (105 or 210 g a.i. ha⁻¹), or saflufenacil (36 g a.i. ha⁻¹) had significantly reduced germination compared to the untreated control (**Table 6**). Adding glyphosate to contact herbicides as a tank mix significantly improved seed germination (9.9%) compared to glyphosate applied alone. In 2013, no adverse effect on seed germination was attributable to glyphosate treatment at either site (**Table 6**).

Similar to seed germination results, glyphosate sprayed alone or tank mix with other contact herbicides, except pyraflufen ethyl (20 g a.i. ha⁻¹) and diquat (208 or 415 g a.i. ha⁻¹) plus glyphosate significantly reduced seedling vigor compared to the untreated control at Saskatoon in 2012 (**Table 7**). On average, the addition of glyphosate to other desiccants as a tank mixture partner significantly reduced seedling vigor compared to their sole application (**Table 7**). Likewise, the glyphosate, glufosinate (600 g a.i. ha⁻¹), saflufenacil (36 or 50 g a.i. ha⁻¹), pyraflufen (100 g a.i. ha⁻¹ or 20 g a.i. ha⁻¹) + glyphosate, glufosinate (300 g a.i. ha⁻¹) + glyphosate (900 g a.i. ha⁻¹), flumioxazin (105 or 210 g a.i. ha⁻¹) + glyphosate, and saflufenacil (36 or 50 g a.i. ha⁻¹) + glyphosate treatments resulted in a significant reduction of seedling vigor compared to the control at Scott in 2012. Overall, the addition of glyphosate to other desiccants as a tank mix significantly reduced seed vigor compared to desiccants applied alone. The high rates of sole application of tank mix herbicides also significantly reduced seedling vigor compared to the lower rate of these herbicides applied alone (**Table 7**).

Conversely, none of the desiccant treatments had a significant effect on seedling vigor in 2013 (**Table 7**). The lack of adverse effects and differences of treatments may have resulted from reduction of glyphosate translocation to lentil seed during desiccation and lower seed moisture content at the time of treatment application. Seed moisture at the time of application was 32 and 35% at Saskatoon and Scott in 2013, respectively, compared to 35 and 40% in 2012.

Milling Characteristics: Dehulling Efficiency, Milling Recovery, and Football Recovery

Dehulling Efficiency (%)

The desiccant by site-year interaction was significant for dehulling efficiency, milling recovery, and football recovery (**Table 3**), and thus these data were analyzed within site-years (**Table 5**). At Saskatoon in 2012, only application of pyraflufen (20 g a.i. ha⁻¹) with glyphosate (900 g a.i. ha⁻¹) significantly reduced dehulling efficiency compared to the

TABLE 2 | Mean monthly temperature and total monthly precipitation during the growing season at Saskatoon and Scott, Saskatchewan, Canada, in 2012 and 2013.

Month	2012		2013	
	Mean monthly temperature (°C)	Total precipitation (mm)	Mean monthly temperature (°C)	Total precipitation (mm)
SASKATOON, SASKATCHEWAN				
May	16.4	108.0	13.0	15.2
June	21.5	121.1	15.2	115.9
July	19.7	80.9	17.4	35.2
Aug.	17.3	48.5	18.9	14.7
Sep.	14.0	0.8	15.2	14.9
SCOTT, SASKATCHEWAN				
May	16.4	50.6	19.9	38.9
June	21.2	164.6	20.0	113.5
July	24.3	56.4	22.7	26.1
Aug.	23.6	51.4	24.5	63.3
Sep.	20.2	24.4	22.1	0.0

Source: <http://climate.weatheroffice.gc.ca>.

TABLE 3 | P-values derived from combined analysis of variance using a mixed model for seed diameter, seed plumpness, seed thickness, seed germination, seed vigor, dehulling efficiency (DE), milling recovery (MR), and football recovery (FR) influenced by desiccation treatment at Saskatoon and Scott, Saskatchewan (SK) in 2012 and 2013.

Source	Seed diameter (d)	Seed thickness (b)	Seed plumpness (b/d)	Seed germination	Seed vigor	DE (%)	MR (%)	FR (%)
Site-year (SY)	0.180	0.163	0.160	0.442	0.190	0.167	0.218	0.016*
Block (Site-year)	0.010*	0.012*	0.013*	0.490	0.004**	0.088	0.044*	0.110
Desiccant (D)	0.152	0.450	0.445	< 0.0001***	< 0.0001***	0.152	0.294	0.028*
SY × D	0.110	0.070	0.058	< 0.0001***	< 0.0001***	0.050*	0.037*	0.050*

*, **, and *** represent significant differences at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively.

TABLE 4 | Mean seed thickness, seed diameter, and seed plumpness of lentil treated by desiccant treatments at Saskatoon and Scott, SK in 2012 and 2013.

Desiccant treatment	Rate(g ai/ae ha ⁻¹)	Seed diameter (mm) (d)	Seed thickness (mm) (b)	Seed plumpness (d/b)
Untreated control	-	4.70	2.71	0.57
Glyphosate	900	4.68	2.69	0.57
Pyraflufen-ethyl	10	4.67	2.66	0.57
Pyraflufen-ethyl	20	4.68	2.62	0.56
Glufosinate	300	4.65	2.65	0.57
Glufosinate	600	4.66	2.67	0.57
Flumioxazin	105	4.67	2.68	0.57
Flumioxazin	210	4.68	2.76	0.59
Saflufenacil	36	4.66	2.67	0.57
Saflufenacil	50	4.69	2.68	0.57
Diquat	208	4.64	2.63	0.56
Diquat	415	4.66	2.62	0.56
Pyraflufen-ethyl + glyphosate	10 + 900	4.69	2.59	0.55
Pyraflufen-ethyl + glyphosate	20 + 900	4.68	2.64	0.56
Glufosinate + glyphosate	300 + 900	4.67	2.60	0.56
Glufosinate + glyphosate	600 + 900	4.64	2.68	0.57
Flumioxazin + glyphosate	105 + 900	4.66	2.69	0.58
Flumioxazin + glyphosate	210 + 900	4.67	2.63	0.56
Saflufenacil + glyphosate	36 + 900	4.66	2.68	0.57
Saflufenacil + glyphosate	50 + 900	4.68	2.62	0.56
Diquat + glyphosate	208 + 900	4.68	2.73	0.57
Diquat + glyphosate	415 + 900	4.65	2.67	0.57
LSD		ns	ns	ns
CONTRASTS				
Untreated control vs. TM ^T and glyphosate		0.03*	0.01	0.05
Glyphosate vs. TM ^T + glyphosate		-0.03*	0.01	0.06
TM ^T vs. TM ^T + glyphosate		0.00	0.00	-0.02
TM ^T (low rate) vs. TM ^T (high rate)		0.02**	0.00	0.00

Contrast statements indicate differences in means between treatments.

TM^T, herbicides used as tank-mix; ns denotes non-significant at $P < 0.05$. * and ** represent significant differences at $P < 0.05$ and $P < 0.01$, respectively. LSD denotes Fisher's least significant differences value at 0.05 level of probability.

control. Application of diquat (415 g a.i. ha⁻¹) increased dehulling efficiency by 5.6% over pyraflufen (20 g a.i. ha⁻¹) with glyphosate (Table 8). The contrast results show application of glyphosate in tank mixtures significantly lowered dehulling efficiency compared to sole application of all contact herbicides.

At Scott in 2012, most desiccant treatments exhibited better or comparable dehulling efficiencies (%) compared to the untreated

control; the only exception was treatment with saflufenacil (36 g a.i. ha⁻¹), which led to significantly reduced dehulling efficiency. Application of the high rate of diquat (415 g a.i. ha⁻¹) increased dehulling efficiency by 4.3% over application of saflufenacil (36 g a.i. ha⁻¹) alone. None of treatments applied at either rate had a significant impact on dehulling efficiency percentages at either Saskatoon or Scott in 2013 (Table 8).

TABLE 5 | F-values from analysis of variance (ANOVA) for seed germination, seed vigor, dehulling efficiency, milling recovery, and football recovery evaluated at Saskatoon and Scott, SK in 2012 and 2013.

Year	Location	Sources of variation	Df	F-values				
				Seed germination	Seed Vigor	Dehulling efficiency (%)	Milling recovery (%)	Football recovery (%)
2012	Saskatoon	Desiccants	21	28.02**	4.07***	1.97*	1.82*	1.73*
	Scott	Desiccants	21	29.2***	7.68***	2.14**	1.84*	0.82 ^{ns}
2013	Saskatoon	Desiccants	21	1.22 ^{ns}	0.78 ^{ns}	1.23 ^{ns}	1.09 ^{ns}	3.47**
	Scott	Desiccants	21	1.31 ^{ns}	1.10 ^{ns}	0.92 ^{ns}	1.45 ^{ns}	1.91*

*, **, and *** represent significant differences at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively. ^{ns} represents non-significant. Df denotes degree of freedom.

TABLE 6 | Means comparison of seed germination (%) of lentil influenced by desiccants at Saskatoon and Scott, SK in 2012 and 2013.

Treatment	Seed germination (%)				
	Rate (g ai/ae ha ⁻¹)	Saskatoon		Scott	
		2012	2013	2012	2013
Untreated control	-	90.0 ^{cd}	95.3	89.5 ^{a-d}	90.7
Glyphosate	900	66.3 ^{hi}	87.5	69.3 ^l	91.0
Pyrflufen-ethyl	10	92.3 ^{a-c}	94.8	91.3 ^{a-c}	89.5
Pyrflufen-ethyl	20	92.0 ^{bc}	94.3	90.3 ^{a-c}	92.2
Glufosinate	300	86.0 ^{a-f}	92.0	79.3 ^{g-i}	88.8
Glufosinate	600	75.8 ^g	90.8	78.5 ^{h-j}	89.7
Flumioxazin	105	95.3 ^a	96.8	91.3 ^{a-c}	92.5
Flumioxazin	210	92.8 ^{a-c}	94.5	93.5 ^a	87.5
Saflufenacil	36	87.0 ^{de}	95.5	88.0 ^{a-f}	91.7
Saflufenacil	50	87.5 ^{de}	95.3	88.8 ^{a-e}	87.2
Diquat	208	91.5 ^{a-e}	95.5	94.0 ^a	90.0
Diquat	415	92.5 ^b	94.0	93.0 ^{a-d}	90.5
Pyrflufen-ethyl + glyphosate	10 + 900	65.3 ^{hi}	93.0	71.0 ^{g-l}	90.5
Pyrflufen-ethyl + glyphosate	20 + 900	71.3 ^{g-i}	93.8	77.5 ^{f-i}	90.2
Glufosinate + glyphosate	300 + 900	73.5 ^{gh}	95.8	78.5 ^{e-j}	89.5
Glufosinate + glyphosate	600 + 900	73.8 ^{g-i}	94.5	80.0 ^{c-i}	88.5
Flumioxazin + glyphosate	105 + 900	64.0 ⁱ	91.3	69.3 ⁱ	90.7
Flumioxazin + glyphosate	210 + 900	66.8 ^{hi}	94.8	73.5 ^{g-i}	90.7
Saflufenacil + glyphosate	36 + 900	74.5 ^{f-i}	92.5	72.3 ^{hi}	91.5
Saflufenacil + glyphosate	50 + 900	78.8 ^{fg}	92.8	76.3 ^{e-l}	89.2
Diquat + glyphosate	208 + 900	86.8 ^{d-f}	96.0	83.8 ^{a-g}	84.7
Diquat + glyphosate	415 + 900	86.5 ^{d-f}	95.8	82.5 ^{b-h}	89.5
LSD		6.70	ns	7.85	ns
CONTRASTS					
Untreated control vs. TM ^T and glyphosate		12.25***	2.14	10.90***	0.45
Glyphosate vs. TM ^T + glyphosate		10.87***	1.95	9.97***	0.50
TM ^T vs. TM ^T + glyphosate		-2.88***	-1.33	-1.95	0.48
TM ^T (low rate) vs. TM ^T (high rate)		5.10**	-0.40	4.05*	1.45

Contrast statements indicate differences in mean between desiccant treatments. Means followed by the same letter within a column are not significantly different ($p < 0.05$). TM^T represents herbicides used as tank-mix; ns denotes non-significant at $P < 0.05$. *, **, and *** represent significant differences at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively. LSD denotes Fisher's least significant differences value at 0.05 level of probability.

Milling Recovery (%)

At Saskatoon in 2012, most desiccant treatments had no significant effect on milling recovery. The glufosinate (300 g a.i. ha⁻¹) and Pyrflufen (20 g a.i. ha⁻¹) glyphosate (900 g a.e.

ha⁻¹) treatment significantly reduced milling recovery compared to control. While diquat (415 g a.i. ha⁻¹) increased milling recovery by 5.0% compared to the Pyrflufen (20 g a.i. ha⁻¹) with glyphosate treatment (Table 9). On average, adding glyphosate

TABLE 7 | Means comparison of seed vigor (%) of lentil influenced by desiccants at Saskatoon and Scott, SK in 2012 and 2013.

Treatment	Rate	Seed vigor (%)			
		Saskatoon		Scott	
	(g ai/ae ha ⁻¹)	2012	2013	2012	2013
Untreated control	-	78.0 ^{a-c}	89.5	87.7 ^{a-c}	83.0
Glyphosate	900	65.0 ^{d-h}	87.5	68.7 ^{hi}	82.2
Pyrflufen-ethyl	10	75.3 ^{a-d}	86.7	82.2 ^{b-e}	85.0
Pyrflufen-ethyl	20	75.3 ^{a-d}	89.5	86.2 ^{a-d}	82.0
Glufosinate	300	78.5 ^{ab}	80.5	80 ^{c-f}	82.0
Glufosinate	600	70.0 ^{b-f}	84.2	72.7 ^{f-i}	81.8
Flumioxazin	105	86.3 ^a	90.7	87.0 ^{a-c}	82.7
Flumioxazin	210	76.0 ^{a-d}	91.5	89.5 ^{ab}	80.7
Saflufenacil	36	70.3 ^{b-e}	90.0	73.0 ^{f-i}	78.0
Saflufenacil	50	61.8 ^{e-h}	87.2	76.5 ^{e-h}	84.7
Diquat	208	77.8 ^{a-c}	89.6	91.0 ^a	84.7
Diquat	415	78.3 ^{a-c}	88.0	89.0 ^{ab}	85.0
Pyrflufen-ethyl + glyphosate	10 + 900	61.5 ^{e-h}	86.8	74.7 ^{e-i}	85.3
Pyrflufen-ethyl + glyphosate	20 + 900	67.0 ^{c-g}	86.5	69.0 ^{hi}	78.7
Glufosinate + glyphosate	300 + 900	66.2 ^{a-d}	88.5	71.7 ^{g-i}	81.3
Glufosinate + glyphosate	600 + 900	66.3 ^{d-h}	88.7	75.2 ^{e-i}	82.5
Flumioxazin + glyphosate	105 + 900	55.5 ^h	86.5	69.5 ^{hi}	84.5
Flumioxazin + glyphosate	210 + 900	57.3 ^{gh}	83.5	68.5 ⁱ	80.5
Saflufenacil + glyphosate	36 + 900	65.0 ^{d-h}	86.5	70.5 ^{hi}	76.5
Saflufenacil + glyphosate	50 + 900	58.8 ^{f-h}	86.0	72.7 ^{f-i}	82.0
Diquat + glyphosate	208 + 900	72.5 ^{b-e}	83.5	82.5 ^{b-e}	82.5
Diquat + glyphosate	415 + 900	71.8 ^{b-e}	87.5	78.7 ^{d-g}	81.7
LSD		11.37	ns	7.81	ns
CONTRASTS					
Untreated control vs. TM ^T and glyphosate		8.30**	1.98	11.63***	0.45
Glyphosate vs. TM ^T + glyphosate		6.88*	1.88	10.97***	0.65
TM ^T vs. TM ^T + glyphosate		2.20	0.13	-1.22	0.67
TM ^T (low rate) vs. TM ^T (high rate)		1.40	0.90	6.20***	2.25

Contrast statements indicate differences in means between lentil desiccant treatments. Means followed by the same letter within a column are not significantly different ($p < 0.05$). TM^T represents herbicides used as tank-mix; ns denotes non-significant at $P < 0.05$. *, **, and *** represent significant differences at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively. LSD denotes Fisher's least significant differences value at 0.05 level of probability.

to desiccants reduced milling recovery by 1.2% compared to sole application of glyphosate. The low application rate of other contact herbicides increased milling recovery (1.4%) compared to the corresponding high rates.

At Scott in 2012, only plots treated with saflufenacil (36 g a.i. ha⁻¹) had significantly lower milling recovery yield compared to the diquat treated plots. Application of the high rate of diquat (415 g a.i. ha⁻¹) increased milling recovery by 4.6% over the low rate of saflufenacil application. Contrast comparisons showed that neither sole application nor herbicide mixes with glyphosate significantly affected milling recovery. No differences in milling recovery were observed between high and low rates of herbicide application treatments (Table 9).

No desiccant treatments affected milling recovery (%) of lentil at Saskatoon or Scott in 2013 (Table 9). Milling recovery was also unaffected by either the addition of contact herbicides to glyphosate or application rate of these herbicides.

Football Recovery (%)

Football recovery yield was influenced by growing environment. At Saskatoon in 2012, most desiccant treatments had a marginal effect on football recovery (Table 10); the exception was significantly reduced recovery for the saflufenacil (50 g a.i. ha⁻¹) + glyphosate (900 g a.e. ha⁻¹) treatment compared to the control. Application of diquat (207 g a.i. ha⁻¹) increased football recovery by 9.5% compared to saflufenacil (50 g a.i. ha⁻¹) + glyphosate (900 g a.e. ha⁻¹). On average, adding glyphosate to other desiccants did not reduce football recovery compared to their sole application at either application rate. At Saskatoon in 2013, application of flumioxazin (210 g a.i. ha⁻¹), diquat (207 g a.i. ha⁻¹), and saflufenacil (36 g a.i. ha⁻¹) with glyphosate significantly improved football recovery compared to the untreated control. At Scott in 2012, none of desiccants applied alone or as a tank mixture with glyphosate significantly reduced football recovery compared to the control. At Scott

TABLE 8 | Means comparison of dehulling efficiency (%) of lentil treated with desiccants at Saskatoon and Scott, SK in 2012 and 2013.

Treatment	Rate (g ai/ae ha ⁻¹)	Dehulling efficiency (%)			
		Saskatoon		Scott	
		2012	2013	2012	2013
Untreated control	-	92.9 ^{a-d}	97.5	94.1 ^{a-c}	96.7
Glyphosate	900	92.3 ^{a-e}	96.3	94.2 ^{a-c}	96.5
Pyraflufen-ethyl	10	93.0 ^{a-d}	97.7	93.7 ^{bc}	95.1
Pyraflufen-ethyl	20	91.8 ^{a-e}	98.2	93.6 ^{bc}	95.3
Glufosinate	300	92.9 ^{a-e}	97.6	93.7 ^{bc}	95.1
Glufosinate	600	92.2 ^{a-e}	97.6	94.8 ^{ab}	96.4
Flumioxazin	105	93.3 ^{a-c}	97.9	93.8 ^{a-c}	96.2
Flumioxazin	210	92.2 ^{a-e}	97.7	94.1 ^{a-c}	96.4
Saflufenacil	36	92.1 ^{a-e}	97.6	91.1 ^d	94.7
Saflufenacil	50	90.4 ^{d-f}	98.1	92.6 ^{cd}	96.5
Diquat	208	93.0 ^{a-c}	97.9	95.0 ^{ab}	95.6
Diquat	415	94.1 ^a	97.8	95.4 ^a	95.6
Pyraflufen-ethyl + glyphosate	10 + 900	90.7 ^{c-f}	97.1	94.2 ^{a-c}	95.6
Pyraflufen-ethyl + glyphosate	20 + 900	88.5 ^f	97.1	94.4 ^{ab}	96.6
Glufosinate + glyphosate	300 + 900	90.3 ^{ef}	97.5	94.5 ^{ab}	96.4
Glufosinate + glyphosate	600 + 900	91.2 ^{b-e}	98.0	94.4 ^{ab}	96.1
Flumioxazin + glyphosate	105 + 900	92.8 ^{a-e}	97.9	93.4 ^{bc}	96.6
Flumioxazin + glyphosate	210 + 900	92.9 ^{a-d}	97.9	93.9 ^{a-c}	96.1
Saflufenacil + glyphosate	36 + 900	91.6 ^{a-e}	97.5	93.7 ^{a-c}	95.2
Saflufenacil + glyphosate	50 + 900	92.8 ^{a-e}	98.1	94.1 ^{a-c}	96.2
Diquat + glyphosate	208 + 900	93.7 ^{ab}	97.5	95.0 ^{ab}	96.2
Diquat + glyphosate	415 + 900	92.4 ^{a-e}	98.2	94.5 ^{ab}	96.0
LSD		2.62	ns	1.70	ns
CONTRASTS					
Untreated control vs. TM ^T and glyphosate		1.36	-0.07	0.06	0.71
Glyphosate vs. TM ^T + glyphosate		1.48	-0.03	-0.01	0.77
TM ^T vs. TM ^T + glyphosate		-1.08*	-0.37	0.48	0.02
TM ^T (low rate) vs. TM ^T (high rate)		1.49	-0.21	-0.36	-0.24

Contrast statements indicate differences in mean between desiccant treatments in lentil. Means followed by the same letter within a column are not significantly different ($p < 0.05$). TM represents herbicides used as tank-mix; ns denotes non-significant at $P < 0.05$. *Represents significant differences at $P < 0.05$, LSD denotes Fisher's least significant differences value at 0.05 level of probability.

in 2013, only application of glufosinate (300 g a.i. ha⁻¹) with glyphosate and saflufenacil (50 g a.i. ha⁻¹) with glyphosate significantly decreased the football recovery compared to the untreated control. Overall, glyphosate tank mixes with other contact herbicides or these herbicides applied in higher doses had no significant impact on football recovery in any site year (Table 10).

Correlation among Lentil Seed Morphology Traits, Glyphosate Residue, and Milling Characteristics

Pearson correlation coefficients (r) were calculated for glyphosate residue content in seeds with other parameters measured in this study. Data were averaged for three replications and combined over both site-years (Saskatoon and Scott) in 2012 and 2013 for correlation analysis. Positive and high correlation was observed between seed thickness and seed plumpness ($r = 0.97$, $p <$

0.001), seed diameter ($r = 0.62$, $p < 0.001$), and glyphosate residue content ($r = 0.51$, $p < 0.001$). Seed thickness was negatively correlated with seed germination, ($r = -0.46$, $p < 0.001$), seedling vigor ($r = -0.26$, $p < 0.01$), dehulling efficiency ($r = -0.33$, $p < 0.01$), and milling recovery ($r = -0.63$, $p < 0.001$; Table 11). Seed diameter was positively and significantly correlated with seed plumpness ($r = 0.51$, $p < 0.001$) and glyphosate residue content ($r = 0.53$, $p < 0.001$) and negatively correlated with other biological traits (Table 11). Likewise, seed plumpness was only positively correlated with glyphosate residue content ($r = 0.48$, $p < 0.001$). Seed germination ($r = -0.84$, $p < 0.001$) and seedling vigor ($r = -0.62$, $p < 0.001$) were negatively and significantly correlated with glyphosate residues. Percent seed germination was positively correlated with seed vigor ($r = 0.75$, $p < 0.001$), dehulling efficiency ($r = 0.68$, $p < 0.001$), and milling recovery ($r = -0.62$, $p < 0.001$).

For milling characteristics, dehulling efficiency was significantly but negatively correlated with glyphosate residue

TABLE 9 | Means comparison of milling recovery (%) of lentil treated with desiccants at Saskatoon and Scott, SK in 2012 and 2013.

Treatment	Rate (g ai/ae ha ⁻¹)	Milling Recovery (%)			
		Saskatoon		Scott	
		2012	2013	2012	2013
Untreated control	-	81.0 ^{a-d}	83.6	79.3 ^{ab}	84.9
Glyphosate	900	80.2 ^{a-e}	82.9	79.4 ^{ab}	84.3
Pyrflufen-ethyl	10	80.7 ^{a-e}	83.8	78.9 ^{ab}	83.9
Pyrflufen-ethyl	20	79.7 ^{a-f}	84.2	79.1 ^{ab}	83.2
Glufosinate	300	78.2 ^{ef}	83.3	79.2 ^{ab}	83.6
Glufosinate	600	80.6 ^{a-e}	83.7	79.6 ^a	84.5
Flumioxazin	105	81.3 ^{a-c}	84.0	79.5 ^{ab}	84.0
Flumioxazin	210	80.5 ^{a-e}	83.7	79.2 ^{ab}	82.3
Saflufenacil	36	79.7 ^{a-e}	83.2	76.2 ^b	82.8
Saflufenacil	50	79.5 ^{b-f}	83.6	78.4 ^{ab}	84.3
Diquat	208	80.9 ^{a-d}	84.5	80.1 ^a	83.5
Diquat	415	82.2 ^a	83.6	80.8 ^a	83.2
Pyrflufen-ethyl + glyphosate	10 + 900	78.8 ^{c-f}	82.9	78.7 ^{ab}	83.8
Pyrflufen-ethyl + glyphosate	20 + 900	77.2 ^f	83.4	79.1 ^{ab}	85.3
Glufosinate + glyphosate	300 + 900	78.5 ^{d-f}	82.5	79.5 ^{ab}	84.8
Glufosinate + glyphosate	600 + 900	79.3 ^{b-f}	83.5	79.4 ^{ab}	84.2
Flumioxazin + glyphosate	105 + 900	80.8 ^{a-e}	83.9	78.4 ^{ab}	86.0
Flumioxazin + glyphosate	210 + 900	81.0 ^{a-d}	83.4	79.4 ^{ab}	84.0
Saflufenacil + glyphosate	36 + 900	79.6 ^{b-f}	83.7	78.7 ^{ab}	82.9
Saflufenacil + glyphosate	50 + 900	80.4 ^{a-e}	83.3	79.2 ^{ab}	83.8
Diquat + glyphosate	208 + 900	81.8 ^a	83.2	79.8 ^a	83.8
Diquat + glyphosate	415 + 900	78.9 ^{c-f}	83.7	79.8 ^a	83.7
LSD		2.58	ns	1.76	ns
CONTRASTS					
Untreated control vs. TM ^T and glyphosate		1.43	0.41	0.09	0.57
Glyphosate vs. TM ^T + glyphosate		1.55	0.49	0.01	0.73
TM ^T vs. TM ^T + glyphosate		-1.19**	-0.42	0.34	0.49
TM ^T (low rate) vs. TM ^T (high rate)		1.35*	0.15	-0.32	-0.43

Contrast statements indicate differences in mean between desiccant treatments. Means followed by the same letter within a column are not significantly different ($p < 0.05$). TM^T represents herbicides used as tank-mix; ns denotes non-significant at $P < 0.05$. * and ** represents significant differences at $P < 0.05$ and $P < 0.01$, respectively, LSD denotes Fisher's least significant differences value at 0.05 level of probability.

($r = -0.55$, $p < 0.001$) and football recovery ($r = -0.64$, $p < 0.001$). Milling recovery also strongly but negatively correlated with glyphosate residues ($r = -0.62$, $p < 0.001$). In contrast, football recovery was positively correlated with glyphosate residues ($r = 0.30$, $p < 0.01$; Table 11).

DISCUSSION

Uniform and early seed maturity is critical to produce high quality lentil on the Canadian Prairies. Extreme growing conditions combined with the heterogeneity of soil, precipitation patterns, and the indeterminate growth habit of lentil plants often result in uneven maturation of the crop. Crop desiccants are used as pre-harvest aids to rapidly dry vegetative and reproductive plant tissues, including seeds, without affecting seed yield and quality (Ratnayake and Shaw, 1992; Soltani et al., 2013). Most lentil growers in Western Canada use herbicidal desiccants to overcome challenges of heterogeneous maturity of

the crop. However, some desiccants directly impact physiological aspects of different crop species, such as mean seed weight, seed germination, and dehulling efficiency (Darwent et al., 2000; Bruce, 2008).

The current study determined the impact of the use of contact herbicides as harvest aids as applied alone or in combination with glyphosate on selected physical, physiological, and processing characteristics of lentils. None of the desiccants applied alone or in tank mixes with glyphosate adversely affected seed physical qualities, including seed diameter, thickness, and plumpness. These results are similar to Ratnayake and Shaw (1992) who report that pre-harvest use of glufosinate, glyphosate, or paraquat had no significant adverse effect on seed yield or quality in soybean when applied at the full maturity stage. Similarly, Zhang et al. (2016) and McNaughton et al. (2015) observed no reduction in yield or thousand seed weight when desiccants were applied to lentil and dry bean, respectively. Wilson and Smith (2002) report that glufosinate, paraquat, and diquat applied

TABLE 10 | Means comparison of football recovery (%) of lentil influenced by application of desiccants at Saskatoon and Scott, SK in 2012 and 2013.

Treatment	Rate (g ai/ae ha ⁻¹)	Football recovery (%)			
		Saskatoon		Scott	
		2012	2013	2012	2013
Untreated control	-	48.1 ^{a-d}	10.1 ^{c-e}	13.7	20.1 ^{a-c}
Glyphosate	900	46.5 ^{b-e}	9.7 ^{c-e}	12.2	17.9 ^{a-e}
Pyraflufen-ethyl	10	49.9 ^{ab}	10.9 ^{c-e}	12.4	20.6 ^{ab}
Pyraflufen-ethyl	20	47.2 ^{b-e}	11.5 ^{b-d}	14.6	16.9 ^{b-f}
Glufosinate	300	49.9 ^{ab}	10.9 ^{c-e}	12.7	21.4 ^a
Glufosinate	600	45.5 ^{b-e}	10.3 ^{c-e}	14.0	16.4 ^{c-f}
Flumioxazin	105	47.8 ^{a-d}	10.4 ^{c-e}	12.2	18.4 ^{a-d}
Flumioxazin	210	46.9 ^{b-e}	15.1 ^a	12.8	18.0 ^{a-e}
Saflufenacil	36	43.9 ^{de}	9.2 ^{de}	11.6	18.0 ^{a-e}
Saflufenacil	50	47.2 ^{b-e}	9.6 ^{c-e}	12.7	16.8 ^{b-f}
Diquat	208	52.2 ^a	13.4 ^{ab}	12.3	16.6 ^{c-f}
Diquat	415	50.1 ^{ab}	11.6 ^{bc}	15.0	16.2 ^{d-f}
Pyraflufen-ethyl + glyphosate	10 + 900	45.5 ^{b-e}	9.3 ^{c-e}	11.8	18.7 ^{ad}
Pyraflufen-ethyl + glyphosate	20 + 900	44.9 ^{c-e}	10.8 ^{c-e}	12.0	17.9 ^{a-e}
Glufosinate + glyphosate	300 + 900	49.6 ^{a-c}	9.0 ^e	12.4	13.4 ^f
Glufosinate + glyphosate	600 + 900	48.5 ^{a-d}	10.1 ^{c-e}	12.5	18.9 ^{a-d}
Flumioxazin + glyphosate	105 + 900	48.4 ^{a-d}	9.4 ^{c-e}	12.3	17.4 ^{b-e}
Flumioxazin + glyphosate	210 + 900	47.6 ^{a-d}	10.3 ^{c-e}	15.0	19.6 ^{a-d}
Saflufenacil + glyphosate	36 + 900	47.4 ^{b-e}	13.4 ^{ab}	13.6	19.1 ^{a-d}
Saflufenacil + glyphosate	50 + 900	42.7 ^e	9.7 ^{c-e}	11.2	14.5 ^{ef}
Diquat + glyphosate	208 + 900	48.6 ^{a-c}	9.5 ^{c-e}	12.0	16.6 ^{c-f}
Diquat + glyphosate	415 + 900	47.2 ^{b-e}	11.7 ^{bc}	14.4	17.5 ^{b-e}
LSD		4.70	2.36	ns	3.79
CONTRASTS					
Untreated control vs. TM ^T and glyphosate		0.53	-0.08	0.98	2.11
Glyphosate vs. TM ^T + glyphosate		0.62	0.16	1.04	2.06
TM ^T vs. TM ^T + glyphosate		0.30	-0.74	0.06	0.72
TM ^T (low rate) vs. TM ^T (high rate)		-0.22	0.09	0.26	1.21

Contrast statements indicate mean differences between desiccant treatments in lentil. Means followed by the same letter within a column are not significantly different ($p < 0.05$). TM^T represents herbicides used as tank-mix; ns denotes non-significant at $P < 0.05$. LSD denotes Fisher's least significant differences value at 0.05 level of probability.

TABLE 11 | Correlation coefficients among lentil seed morphology traits, glyphosate residue, and milling characteristics of lentil.

Traits	STH	SD	SP	SG	SV	DE	FR	MR	GR
STH	-								
SD	0.62 ^{***}	-							
SP	0.97 ^{***}	0.51 ^{***}	-						
SG	-0.46 ^{***}	-0.46 ^{***}	-0.42 ^{***}	-					
SV	-0.26 ^{**}	-0.39 ^{***}	-0.23 [*]	0.75 ^{***}	-				
DE	-0.33 ^{**}	-0.45 ^{***}	-0.29 ^{***}	0.68 ^{***}	0.75 ^{***}	-			
FR	-0.16 ^{ns}	0.31 ^{***}	-0.21 [*]	-0.49 ^{***}	0.68 ^{***}	-0.64 ^{***}	-		
MR	-0.63 ^{***}	-0.68 ^{***}	-0.57 ^{***}	0.62 ^{***}	0.49 ^{***}	0.76 ^{***}	-0.23 [*]	-	
GR	0.51 ^{***}	0.53 ^{***}	0.48 ^{***}	-0.84 ^{***}	-0.62 ^{***}	-0.55 ^{***}	0.30 ^{**}	-0.62 ^{***}	-

STH, Seed thickness; SD, Seed diameter; SP, Seed plumpness; SG, Seed germination; SV, Seed vigor; DE, Dehulling efficiency; FR, Football recovery; MR, Milling recovery; GR, glyphosate residue. *, **, and *** indicate significance at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively; ns, non-significant.

as desiccants to common bean accelerated seed maturity and desiccation. Glyphosate applied as a desiccant reduced pod length and seed weight when used as a harvest aid in cowpea (*Vigna unguiculata* L.) (Cedeira et al., 1985). Many studies report reduced soybean yield and seed quality as a result of the application of desiccant prior to crop maturity (Azlin and Mcwhorter, 1981; Cerkauskas et al., 1982; Boudreaux and Griffin, 2011). On the other hand, Soltani et al. (2013) observed that the addition of diquat, glufosinate, carfentrazone, flumioxazin, or saflufenacil to glyphosate improved the drying of dry bean foliage and yield. The variability among results may be related to the timing of desiccant application. The application of desiccants before physiological maturity can inhibit photosynthesis or, because lentil is an indeterminate crop where bottom pods matured before upper canopy's pods, may damage immature seeds.

In 2012, the application of glyphosate alone or as a tank mix with other herbicides (except diquat + glyphosate) significantly reduced germination percentage in lentil seeds compared to the untreated control. These results are similar to those reported by Yenish and Young (2000), who found that pre-harvest glyphosate-treated wheat had a 2–46% lower seed germination than the control. Similarly, Hampton and Hebblethwaite (1982) found that pre-harvest application of glyphosate significantly lowered ryegrass (*Lolium perenne* L.) seed germination due to production of abnormal seedlings. They also reported that germination percentage of the glyphosate-treated seeds decreased with storage. Bennett and Shaw (2000a) report that sodium chlorate plus glyphosate or paraquat applied as a pre-harvest aid to sicklepod (*Senna obtusifolia* L.) 14 days before harvest significantly reduced shoot growth and seed germination.

The reduced seed germination of glyphosate-treated plants may be caused by translocation of glyphosate to the maturing seeds and embryo, the major sink during the maturation process. Glyphosate inhibits the shikimate acid pathway for synthesis of branched aromatic amino acids, such as phenylalanine, tyrosine, and tryptophan (Vivancos et al., 2011). Tryptophan is a direct precursor of indole-3-acetic acid (IAA), which affects coleoptile elongation and shoot and root initiation (Taiz and Zeiger, 1998). Clay and Griffin (2000) suggest that use of glyphosate as a desiccant during the plant maturation phase may affect the level of IAA, the main endogenous auxin, thereby causing inhibition of germination and growth. Unlike glyphosate, diquat applied alone or with glyphosate, pyraflufen, glufosinate (300 g a.i. ha⁻¹), flumioxazin, and saflufenacil applied alone did not affect seed germination in the present study. Ratnayake and Shaw (1992) report similar results. Whigham and Stoller (1979) also observed paraquat applied as a harvest aid had no effect on germination percentage of soybean. All of these are contact herbicides with limited phloem mobility, and therefore low levels of these compounds would reach the seed.

Our study showed no adverse effect of glyphosate or any contact herbicides on seed germination of lentil at either site in 2013. This might be due to lower moisture content in seeds due to reduced rainfall at the time of application (Table 2). Low rainfall hastened the dry down process of lentil crops in 2013. Zhang et al. (2016) reported glyphosate residue in lentil

seeds content <2 ppm from Saskatoon and Scott when they had 32 and 35% moisture content in 2013 compared with 35 and 40% in 2012, respectively. Lower moisture content in seed harvested in 2013 might have reduced translocation of glyphosate in seeds. The application of glyphosate as a desiccant in lentil (Zhang et al., 2017) and common bean (McNaughton et al., 2015) prior to 30% seed moisture content, increases its residue to an unacceptable level (>2 ppm) causing reduced seed yield and mean seed weight. Higher translocation of glyphosate into seeds may also reduce seed germination and vigor as developing seeds are major photosynthesis sinks (Zhang et al., 2016). The significant and negative correlation between glyphosate residue and seed germination results support this explanation. Different countries do have different import policies in terms of the amount of glyphosate residue they will accept for import of lentil (Pratt, 2011). The current MRLs for glyphosate in lentil are 2 ppm, 4 ppm and 10 ppm for Canada, Japan, and the European Union (EU), respectively (Zhang et al., 2017).

A significant reduction in lentil seedling vigor was observed at Saskatoon and Scott in 2012 when glyphosate was sprayed alone or as a tank mix with other herbicides but no such differences were observed in 2013. The 2012 results are comparable to those of Hampton and Hebblethwaite (1982), who show that seedling vigor of perennial ryegrass (*Lolium perenne* L.) is reduced when glyphosate is applied as a pre-harvest aid. The pre-harvest application of glyphosate is also reported to cause poor seed germination and reduced seedling vigor when applied at >40% seed moisture content in field pea and soybean (Bennett and Shaw, 2000b; Baig et al., 2003). Our study showed seedling vigor was highly negatively correlated with glyphosate residue in lentil seeds. Adverse effects of pre-harvest glyphosate application on seedling growth and vigor might be related to the reduction and physiological inactivation of mineral nutrients, such as Ca and Mn, due to glyphosate in the seeds (Cakmak et al., 2009). Mineral nutrients can play an important role in seed viability and seedling vigor and establishment, particularly under adverse soil conditions (Welch, 1999).

The growing environment can have a significant effect on dehulling efficiency of lentil (Bruce, 2008). Results from the present study show the effects of desiccants on dehulling efficiency, milling recovery, and football recovery of lentils depend to a certain extent on growing environment and moisture content in seeds at the time of application. We observed that pyraflufen (20 g a.i. ha⁻¹) with glyphosate (900 g a.e. ha⁻¹) and saflufenacil (36 g a.i. ha⁻¹) reduced dehulling efficiency at Saskatoon and Scott in 2012, respectively. This results were concurrent with the studies on the effect of desiccation with saflufenacil in lentil (Zhang et al., 2017) and common bean (McNaughton et al., 2015) who found that dramatically reduced mean seed weight and seed yield if application of saflufenacil was made prior to 30% seed moisture content. The saflufenacil residue in the seed was reported as 0.03 mg kg⁻¹. Adding saflufenacil to glyphosate did not reduced glyphosate residue in lentil compared to glyphosate applied alone, yet they found that tank mixture significantly reduced seed residue content of saflufenacil and improved crop desiccation. Saflufenacil residues present in harvested lentils may be a concern for lentil growers

when it is used as a harvest aid at pre-harvest stages of crop because major lentil importing countries have also set MRLs for saflufenacil (Bryant Christie Inc, 2015).

Reduction of dehulling efficiency due to saflufenacil applied alone or as a tank mix with pyraflufen-ethyl might have occurred because these herbicides belong to the uracil and phenyl pyrazole classes, respectively, which are protoporphyrinogen oxidase (PPO) inhibitors (Grossmann et al., 2010). PPO inhibitors bind sites of Protogen IX in the chloroplast, which causes peroxidation of foliar cell membrane lipids and subsequent rapid loss of membrane integrity and necrosis (Duke et al., 1991; Grossmann et al., 2010). Both saflufenacil and pyraflufen have limited mobility in phloem (Liebl et al., 2008). These herbicides translocate mainly through the xylem, and their slow mobility may result in accumulation in seeds, thereby interfering with the normal chemical composition and alignment of the bonding layer between the lentil seed coat and the cotyledon.

Pre-harvest application of the low rate of pyraflufen with glyphosate or sole application of saflufenacil (36 g a.i. ha⁻¹) resulted in reduced milling recovery at Saskatoon and Scott, respectively, in 2012. Application of diquat with or without glyphosate seemed to improve milling recovery in 2012. In contrast, Bruce (2008) reports that, during dry harvest conditions, no significant differences in milling recovery were evident between swathing and desiccation pre-harvest treatments. In our study, differential effects of the pre-harvest treatments on football recovery percentage were observed in both years but only in Saskatoon. Similar to the dehulling efficiency, saflufenacil (50 g a.i. ha⁻¹) combined with glyphosate resulted in the lowest football recovery at Saskatoon in 2012. Different responses of crop harvest aids among years (2012 vs. 2013) and sites might have been the result of wet field conditions during the harvesting period and differences in moisture content and glyphosate residue in seeds. Our study shows that milling parameters, particularly dehulling efficiency and milling recovery, are inversely related to glyphosate residue in seeds. Zhang et al. (2016) report that seed moisture content in seeds at the time of application can strongly impact glyphosate translocation to seeds. They note that, irrespective of desiccation treatments, high moisture content in seeds at the time of application results in accumulation of high (>2.0 ppm) of glyphosate residues compared to lower moisture in seeds while they desiccate lentil crops. In most cases (two exceptions), however, desiccant treatments had no significant impact on football recovery at Scott in either year. These results are comparable with those of Bruce (2008), who report no differences in football recovery between swathing and desiccation with diquat when treatments were applied at a later stage of plant maturity; however, early application of diquat decreased football recovery. He suggests that desiccation by diquat caused lentil seeds to separate at the cotyledons more easily compared to swathing followed by natural drying. Swathing allows biological processes related to cotyledon binding to continue for a given period, whereas desiccation may cease the processes rapidly and therefore make the seeds more brittle.

The inconsistency of some of the results in our study in relation to seed biological qualities and milling recovery

over environment indicate that further research over many environments may be required to determine the consistency of treatment effects and their economic impact. Moisture content of seeds is a key cause of translocation of glyphosate residue to seeds, which can result in adverse effects on both seed biological and milling qualities. Therefore, glyphosate is recommended for use as a desiccant in lentil once the seed moisture is 30% or less (Saskatchewan Ministry of Agriculture, 2016). Future research could focus on the relationship between moisture content of seeds and their other milling and post-harvest qualities. Environmental variation and the nature of herbicide sensitivity of lentil crops may result in different impacts on seed quality; this has been demonstrated in differences of sensitivity among crops species to flumioxazin, saflufenacil, and pyraflufen (Ivany, 2005; Soltani et al., 2010) in dry bean.

CONCLUSIONS

Use of glyphosate alone or in tank mix with other contact herbicides as pre-harvest aids in lentil production adversely affected seed germination and seedling vigor particularly if glyphosate is translocated to the seeds. Consistent improvements in milling recovery and dehulling efficiency of lentil were only observed when diquat was used alone or in tank mix with glyphosate, suggesting that lentil growers should consider these desiccant treatments to optimize dry-down of lentil without harming seed quality if they wish to gain premium prices for red lentil based on milling efficiency.

AUTHOR CONTRIBUTIONS

MS conducted the experiment, data collection, data analysis, interpretation, summarized the results and drafting manuscript. CW supervised the field experiments and assisted with editing the manuscript. AV Co-conceptualized and coordinated and directed the project. AV edited, revised and reviewed and drafting manuscript. All authors read and approved the final manuscript.

FUNDING

This research was supported by the Agriculture Development Fund of the Saskatchewan Ministry of Agriculture, the University of Saskatchewan Department of Plant Sciences, Saskatchewan Pulse Growers and the government of Canada through the NSERC Industrial Research Chair in Lentil Genetic Improvement.

ACKNOWLEDGMENTS

We are grateful to Ti Zhang and Gerry Stuber of the Agronomy and Weed Ecology Laboratory for assistance with field work. Special thanks go to Brent Barlow and the Field Crew of the Pulse Research Field Laboratory, University of Saskatchewan for providing technical assistance. We also thank the staff of Discovery Seed Laboratory, Saskatoon, SK, Canada for their technical support.

REFERENCES

- AACC (2000). "International approved methods of the american association of cereal chemists 2000" in *The Association 10th Edn. Methods* (St. Paul, MN).
- Azlin, W. R., and McWhorter, G. C. (1981). Preharvest effects of applying glyphosate to soybeans (*Glycine max*). *Weed Sci.* 29, 123–127.
- Baig, M., Darwent, A., Harker, K., and O'Donovan, J. (2003). Preharvest applications of glyphosate affect emergence and seedling growth of field pea (*Pisum sativum*). *Weed Technol.* 17, 655–665. doi: 10.1614/Wt-02-075
- Bennett, A. C., and Shaw, D. R. (2000a). Effect of pre-harvest desiccants on weed seed production and viability. *Weed Technol.* 14, 530–538. doi: 10.1614/0890-037X(2000)014[0530:EOPDOW]2.0.CO;2
- Bennett, A. C., and Shaw, D. R. (2000b). Effect of pre-harvest desiccants on Group IV *Glycine max* seed viability. *Weed Sci.* 48, 426–430. doi: 10.1614/0043-1745(2000)048[0426:EOPDOG]2.0.CO;2
- Boudreaux, J. M., and Griffin, J. L. (2011). Application timing of harvest aids herbicides affects soybean harvest and yield. *Weed Technol.* 25, 38–43. doi: 10.1614/WT-D-10-00045.1
- Brand, J. (2008). *International Research Review*. Horsham, VIC: Grain Research and Development Corporation.
- Bruce, J. (2008). *The Effect of Pre-Harvest Treatments on the Milling Efficiency of Red Lentil*. MSc Thesis. Department of Plant Sciences. University of Saskatchewan, Saskatoon.
- Bryant Christie Inc. (2015) *Global MRL Database*. Bryant Christie Inc. Available online at: <https://www.globalmrl.com> (Accessed 14 January, 2015).
- Cakmak, I., Yazici, A., Tutus, Y., and Ozturk, L. (2009). Glyphosate reduced seed and leaf concentrations of calcium, manganese, and iron in non-glyphosate resistant soybean. *Europ. J. Agron.* 31, 114–119. doi: 10.1016/j.eja.2009.07.001
- Cedeira, A. L., Cole, A. W., and Luthe, D. H. (1985). Cowpea (*Vigna unguiculata*) seed protein response to glyphosate. *Weed Sci.* 33, 1–6.
- Cerkaskas, R. F., Dhingra, O. D., Sinclair, J. B., and Foor, S. R. (1982). Effect of three desiccant herbicides on soybean (*Glycine max*) seed quality. *Weed Sci.* 30, 484–490.
- CFIA (2012). *Canadian Methods and Procedures for Testing Seeds*. Saskatoon : Canadian Food Inspection Agency.
- Clay, P. A., and Griffin, J. L. (2000). Weed seed production and seedling emergence responses to late season glyphosate application. *Weed Sci.* 48, 481–486. doi: 10.1614/0043-1745(2000)048[0481:WSPASE]2.0.CO;2
- Cobb, A. H., and Reade, J. P. H. (2010). *Herbicide and Plant Physiology, 2nd Edn*. London: Wiley-Blackwell.
- Darwent, A. L., Kirkland, K. J., Townley-Smith, L., Harker, K. N., and Cessna, A. J. (2000). Effect of pre-harvest applications of glyphosate on the drying, yield and quality of canola. *Can. J. Plant Sci.* 80, 433–439. doi: 10.4141/P99-063
- DellaValle, D. M., Vandenberg, A., and Glahn, R. P. (2013). Seed coat removal improves iron bioavailability in cooked lentils: studies using an *in-vitro* digestion/Caco-2 cell culture model. *J. Agric. Food Chem.* 61, 8084–8089. doi: 10.1021/jf4022916
- Duke, S. O., Lydon, J., Becerril, J. M., Sherman, T. D., Lehnen, L. P., and Matsumoto, H. (1991). The protoporphyrinogen oxidase-inhibiting herbicides. *Weed Sci.* 39, 465–473. <http://www.jstor.org/stable/4044980>
- Fedoruk, M. J., Vandenberg, A., and Bett, K. E. (2013). Quantitative trait loci analysis of seed quality characteristics in lentil using single nucleotide polymorphism markers. *Plant Geno.* 6, 1–10. doi: 10.3835/plantgenome2013.05.0012
- Griffin, J. L., Boudreaux, J. M., and Miller, D. K. (2010). Herbicides as harvest aids. *Weed Sci.* 58, 355–358. doi: 10.1614/WS-09-108.1
- Grossmann, K., Niggweg, R., Christiansen, N., Looser, R., and Ehrhardt, T. (2010). The herbicide saflufenacil (kixortm) is a new inhibitor of protoporphyrinogen ix oxidase activity. *Weed Sci.* 58, 1–9. doi: 10.1614/WS-D-09-00004.1
- Hampton, J. G., and Hebblethwaite, P. D. (1982). The preharvest use of glyphosate in the ryegrass seed crop. *Grass Forage Sci.* 37, 243–248. doi: 10.1111/j.1365-2494.1982.tb01602.x
- Hossain, S., Ford, R., McNeil, D., Pittock, C., and Panozzo, J. F. (2010). Development of a selection tool for seed shape and QTL analysis of seed shape with other morphological traits for selective breeding in chickpea (*Cicer arietinum* L.). *Aus. J. Crop Sci.* 4:278. Available online at: http://www.croproj.com/shanoor_4_4_2010_278_288.pdf
- Ivany, J. A. (2005). Response of three potatoes (*Solanum tuberosum*) cultivars to pyraflufen-ethyl used as a desiccant in Canada. *Crop Prot.* 24, 836–841. doi: 10.1016/j.cropro.2004.12.015
- Jood, S., Bichnoi, S., and Sharma, A. (1998). Chemical analysis and physio-chemical properties of chickpea and lentil cultivars. *Nahrung* 42, 71–74. doi: 10.1002/(SICI)1521-3803(199804)42:02<71::AID-FOOD71>3.3.CO;2-U
- Liebl, R., Walter, H., Bowe, S. J., Holt, T. J., and Westberg, D. E. (2008). "A new herbicide for preplant burndown and preemergence dicot weed control," in *Weed Science Society of America*. Chicago, IL: Conference Abstract.
- McNaughton, K. E., Blackshaw, R. E., Waddell, K. A., Gulden, R. H., Sikkema, P. H., and Gillard, C. L. (2015). Effect of five desiccants applied alone and in combination with glyphosate in dry edible bean (*Phaseolus vulgaris* L.). *Can. J. Plant Sci.* 95, 1235–1242. doi: 10.4141/cjps-2014-157
- Pratt, S. (2011). EU rejects lentils for glyphosate residue. Western Producer. Available online at: <http://www.producer.com/2011/04/eu-rejects-lentils-for-glyphosate-residue/> (Accessed 4 February, 2015).
- Ramakrishnaiah, N., and Kurien, P. P. (1983). Variability in the dehulling characteristics of pigeon pea (*Cajanus cajan* L.) cultivars. *J. Food Sci. Technol.* 20, 287–291.
- Ratnayake, S., and Shaw, D. R. (1992). Effects of harvest aid herbicides on sicklepod (*Cassia obtusifolia*) seed yield and quality. *Weed Technol.* 6, 985–989.
- Risula, D. (2014). *Using Harvest Aids and Products in 2014*. Saskatchewan Ministry of Agriculture. Agriview 10, 4–5. Available online at: https://issuu.com/skagriview/docs/38933_gsk_agriview_julyaug2014_for_i.
- SAS (2015). *SAS/STAT® 13.2 User's Guide*. Cary, NC: SAS Institute Inc. Available online at: <https://support.sas.com/documentation/cdl/en/stathpug/67524/PDF/default/stathpug.pdf> (Accessed February 20, 2015).
- Saskatchewan Ministry of Agriculture (2016). *Guide to Crop Protection*. Saskatchewan Ministry of Agriculture. Available online at: <http://publications.gov.sk.ca/documents/20/83821-d281d596-7317-42f1-ac86-898fa5908d18.pdf>
- Saskatchewan Pulse Growers (2015). *Desiccation of Pulse Crops-Strategies for Success*. Available online at: http://proof.saskpulse.com/files/general/150729_Desiccation_Products_for_Pulses.pdf (Accessed July 29, 2015).
- Saxton, A. M. (1998). "A macro for converting mean separation output to letter groupings in Proc Mixed," in *Proceedings 23rd SAS Users Group Intl., SAS Institute* (Cary, NC).
- Singh, U., and Singh, B. (1992). Tropical grain legumes as important human foods. *Econ. Bot.* 46, 310–321. doi: 10.1007/BF02866630
- Soltani, N., Blackshaw, R. E., Gulden, R. H., Gillard, C. L., Shropshire, C., and Sikkema, P. H. (2013). Desiccation in dry edible beans with various herbicides. *Can. J. Plant Sci.* 93, 871–877. doi: 10.4141/cjps2013-061
- Soltani, N., Shropshire, C., and Sikkema, P. H. (2010). Sensitivity of leguminous crops to saflufenacil. *Weed Technol.* 24, 143–146. doi: 10.1614/WT-09-029.1
- Statistics Canada (2016). *Principal Field Crop Areas*. Available online at: <http://www.statcan.gc.ca/daily-quotidien/160629/dq160629a-eng.htm> (Accessed June 29, 2016).
- Taiz, L., and Zeiger, E. (1998). *Auxins. Plant Physiology*. Sunderland, MA: Sinauer Associates.
- Vandenberg, A. (2009). "Post-harvest processing and value addition," in *The Lentil Botany, Production and Uses*, eds W. Erskine and F. J. Muehlbauer (Cambridge, UK: CAB International), 391–424.
- Vivancos, P. D., Driscoll, S. P., Bulman, C. A., Ying, L., Emami, K., Treumann, A., Mauve, C., et al. (2011). Perturbations of amino acid metabolism associated with glyphosate-dependent inhibition of shikimic acid metabolism affect cellular redox homeostasis and alter the abundance of proteins involved in photosynthesis and photorespiration. *Plant Physiol.* 157, 256–268. doi: 10.1104/pp.111.181024
- Wang, N. (2005). Optimization of a laboratory dehulling process for lentil (*Lens culinaris*). *Cereal Chem.* 82, 671–676. doi: 10.1094/CC-82-0671
- Wang, N. (2008). Effect of variety and crude protein content on dehulling quality and on the resulting chemical composition of red lentil (*Lens culinaris*). *J. Sci. Food Agric.* 88, 885–890. doi: 10.1002/jsfa.3165
- Welch, R. M. (1999). "Importance of seed mineral nutrient reserves in crop growth and development," in *Mineral Nutrition of Crops: Fundamental Mechanisms and Implications*, ed Z. Rengel (New York, NY: Food Products Press), 205–226.

- Whigham, D. K., and Stoller, E. W. (1979). Soybean desiccation by paraquat, glyphosate and ametryn to accelerate harvest. *Agron. J.* 71, 630–633. doi: 10.2134/agronj1979.00021962007100040027x
- Wilson, R. G., and Smith, J. A. (2002). Influence of harvest aid herbicides on dry bean (*Phaseolus vulgaris*) desiccation, seed yield, and quality. *Weed Technol.* 16, 109–166. doi: 10.1614/0890-037X(2002)016[0109:IOHAHO]2.0.CO;2
- Xu, B., and Chang, S. K. C. (2009). Phytochemical profiles and health-promoting effects of cool-season food legumes as influenced by thermal processing. *J. Agric. Food Chem.* 57, 10718–10731. doi: 10.1021/jf902594m
- Yenish, J. P., and Young, F. L. (2000). Effect of preharvest glyphosate application on seed and seedling quality of spring wheat (*Triticum aestivum*). *Weed Technol.* 14, 212–217. doi: 10.1614/0890-037X(2000)014[0212:EOPGAO]2.0.CO;2
- Zhang, T., Johnson, E. N., Mueller, T. C., and Willenborg, C. J. (2017). Early application of dry-down herbicides adversely impacts herbicide residues and seed yield in lentil. *Agron. J.* 109, 239–248. doi: 10.2134/agronj2016.07.0419
- Zhang, T., Johnson, E. N., and Willenborg, C. J. (2016). Evaluation of harvest aid herbicides as desiccants in lentil production. *Weed Technol.* 30, 629–638. doi: 10.1614/WT-D-16-00007.1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Subedi, Willenborg and Vandenberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



OPEN ACCESS

Edited by:

Diego Rubiales,
Spanish National Research Council,
Spain

Reviewed by:

Sara Fondevilla,
Spanish National Research Council,
Spain
Marcello Duranti,
University of Milan, Italy
Rebecca Ford,
Griffith University, Australia

*Correspondence:

Karam B. Singh
Karam.Singh@csiro.au

† Present address:

Su Melser,
INSERM U1215, NeuroCentre
Magendie, Group Endocannabinoids
and Neuroadaptation,
Bordeaux, France;
Université de Bordeaux,
NeuroCentre Magendie,
Bordeaux, France

‡ These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 07 October 2016

Accepted: 24 November 2016

Published: 09 December 2016

Citation:

Jimenez-Lopez JC, Melser S,
DeBoer K, Thatcher LF,
Kamphuis LG, Foley RC and
Singh KB (2016) Narrow-Leafed
Lupin (*Lupinus angustifolius*)
 β 1- and β 6-Conglutin Proteins Exhibit
Antifungal Activity, Protecting Plants
against Necrotrophic Pathogen
Induced Damage from *Sclerotinia*
sclerotiorum and *Phytophthora*
nicotianae. *Front. Plant Sci.* 7:1856.
doi: 10.3389/fpls.2016.01856

Narrow-Leafed Lupin (*Lupinus angustifolius*) β 1- and β 6-Conglutin Proteins Exhibit Antifungal Activity, Protecting Plants against Necrotrophic Pathogen Induced Damage from *Sclerotinia sclerotiorum* and *Phytophthora nicotianae*

Jose C. Jimenez-Lopez^{1,2†}, Su Melser^{3†}, Kathleen DeBoer^{1,3}, Louise F. Thatcher³,
Lars G. Kamphuis^{1,3}, Rhonda C. Foley³ and Karam B. Singh^{1,3*}

¹ The Institute of Agriculture, The University of Western Australia, Perth, WA, Australia, ² Department of Biochemistry, Cell and Molecular Biology of Plants, Estacion Experimental del Zaidin, Spanish National Research Council, Granada, Spain,

³ Centre for Environment and Life Sciences, Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation, Floreat, WA, Australia

Vicilins (7S globulins) are seed storage proteins and constitute the main protein family in legume seeds, particularly in narrow-leafed lupin (*Lupinus angustifolius* L.; NLL), where seven vicilin genes, called β 1- to β 7-conglutin have been identified. Vicilins are involved in germination processes supplying amino acids for seedling growth and plant development, as well as in some cases roles in plant defense and protection against pathogens. The roles of NLL β -conglutins in plant defense are unknown. Here the potential role of five NLL β -conglutin family members in protection against necrotrophic fungal pathogens was investigated and it was demonstrated that recombinant purified 6xHis-tagged β 1- and β 6-conglutin proteins exhibited the strongest *in vitro* growth inhibitory activity against a range of necrotrophic fungal pathogens compared to β 2, β 3, and β 4 conglutins. To examine activity *in vivo*, two representative necrotrophic pathogens, the fungus *Sclerotinia sclerotiorum* and oomycete *Phytophthora nicotianae* were used. Transient expression of β 1- and β 6-conglutin proteins in *Nicotiana benthamiana* leaves demonstrated *in vivo* growth suppression of both of these pathogens, resulting in low percentages of hyphal growth and elongation in comparison to control treated leaves. Cellular studies using β 1- and β 6-GFP fusion proteins showed these conglutins localized to the cell surface including plasmodesmata. Analysis of cellular death following *S. sclerotiorum* or *P. nicotianae* revealed both β 1- and β 6-conglutins suppressed pathogen induced cell death *in planta* and prevented pathogen induced suppression of the plant oxidative burst as determined by protein oxidation in infected compared to mock-inoculated leaves.

Keywords: 7S globulins, fungal pathogen, legume, oxidative stress, plant defense, seed storage protein, vicilins

INTRODUCTION

Plants are under constant exposure to potential microbial pathogens. One of the mechanisms they employ to defend themselves is via the production of bioactive antimicrobial proteins (AMPs). In addition to plants, other organisms may produce a diverse array of AMPs for defense purposes and these can confer a high level of antimicrobial activity against competing microorganisms such as bacteria, viruses, protozoa, filamentous fungi and yeasts (Niyonsaba et al., 2009; Tam et al., 2015). In plants AMPs can play a role in constitutive immunity or can be induced upon pathogen attack. Inducible responses can include the expression of pathogen-related (PR) proteins such as the enzymes (1-3)- β -glucanases (PR-2), chitinases (PR-3, -4, -8, and -11), peroxidases (PR-9) and oxalate oxidases (PR-16 and -17; Thatcher et al., 2005; Tam et al., 2015). In addition, it has been proposed that proteins involved in the delivery of storage and energy requirements to plant embryos during germination may also be involved in defense responses (Chrispeels and Raikhel, 1991; Marcus et al., 1999; Gábrišová et al., 2016). For example, members of the following storage protein families 2S albumins, Kunitz proteinase inhibitors, plant lectins and vicilins or vicilin-like proteins (including 7S globulins and β -conglutins; De Souza Candido et al., 2011).

Plant storage proteins can be classified into vegetative storage proteins and seed storage proteins, where the latter can represent a significant proportion of seed composition (Gomes et al., 2014). Storage proteins perform essential roles in plant survival. They provide a source of amino acids that can be mobilized and utilized for maintenance and growth during both seed embryonic developmental, and germination stages (Zienkiewicz et al., 2011; Tan-Wilson and Wilson, 2012; Jimenez-Lopez et al., 2016). These proteins accumulate in cellular storage vacuoles of seeds, nuts, and kernels; stem parenchyma of trees; grains and legumes; and some roots and tubers. The vicilins, also called conglutins in some legume species, constitute a class of proteins abundantly found as reserves in seeds of leguminous and non-leguminous plants, representing as much as 70 to 80% of total protein in the seeds of these plants (Duranti and Gius, 1997). Their structure consists of a trimeric organization, and unlike most plant storage proteins individual subunits with molecular masses typically around 15–70 kDa (Melo et al., 1994), NLL's individual subunits are larger and range from 150 to 170 kDa in size (Argos et al., 1985).

Vicilins appear to play multifunctional roles, acting as an energy source and providing amino acids during the germination process, while in some cases, also being involved in defense responses against fungi and insects (Yunes et al., 1998). This includes for example, vicilins from the legumes *Vigna unguiculata* (cowpea), *V. radiata* (mung bean), *Phaseolus vulgaris* (common bean) and *Canavalia ensiformis* (jack bean; Gomes et al., 1997, 1998; Oliveira et al., 1999; Coda et al., 2008). The insecticidal activity of vicilins relates to their capacity to bind chitinous structures, thereby interfering with insect development, as shown for cowpea and the cowpea seed beetle (*Callosobruchus maculatus*; Sales et al., 2001). This chitin-binding activity can also inhibit yeast and fungal growth (Gomes et al., 1998). The potency of vicilin antifungal activity varies among plant species.

For example, Gomes et al. (Gomes et al., 1998) extracted a vicilin from *V. unguiculata* showing inhibitory activity between 90 and 100% against the yeast *S. cerevisiae*, in addition to interfering with spore germination of the fungi *Fusarium solani*, *F. oxysporum*, *Colletotrichum musae*, *Phytophthora capsici*, *Neurospora crassa* and *Ustilago maydis sporidia*. Vicilin extracted from *V. radiata* seeds showed 65% inhibitory activity against *Candida albicans* (Gomes et al., 1998), whereas vicilin isolated from the non-legume *Malva parviflora* Malva (an annual or perennial herb) showed inhibitory activity against *Phytophthora infestans* (Wang et al., 2001).

Narrow-leafed lupin (*Lupinus angustifolius* L.; NLL) is a recently domesticated important pulse crop, and increasingly popular due to its wide range of agricultural and health benefits (Berger et al., 2013). The NLL grain constitutes an important source of protein for humans and animals with low starch content and free of gluten (reviewed in Foley et al., 2011). In NLL the seed storage proteins are collectively called conglutins and fall into four sub-families called α , β , γ , and δ -conglutins (Foley et al., 2011, 2015). In addition to dissection of lupin-based health benefits, the identification of lupin seed storage proteins playing roles in resistance against pathogens is of interest. Recently, antifungal activity from a multifunctional glyco-oligomer with 210 kDa, mainly composed by BLAD (*banda de Lupinus albus doce*), a 20 kDa polypeptide, a stable intermediary product of β -conglutinin catabolism, was demonstrated and found to exclusively accumulate in the cotyledons of *Lupinus* species (Monteiro et al., 2015).

The recent development of a reference NLL genome assembly (Hane et al., 2016)¹ and extensive RNA expression analysis from various tissues including seeds (Foley et al., 2015; Kamphuis et al., 2015) facilitated the identification of 16 *conglutin* genes, where the β -*conglutin* family was the most abundant, representing 56% of the total seed storage protein RNA expression levels (Foley et al., 2011). The NLL β -conglutin family comprises seven members, namely β 1- to β 7-conglutin (Foley et al., 2011). These β -conglutins share sequence identities ranging from 77.4 to 94.7%, reflected presumably in differential structure-functionality between some of them (Jin et al., 2014), and are highly expressed in the seeds compared to other NLL tissues (Foley et al., 2015).

Pathogenic fungi of lupins, as is the case for many other grain legume crops, cause substantial annual crop losses and are of major economic significance (Rubiales et al., 2015). For example, *Sclerotinia* stem rot, *Rhizoctonia* barepatch, *Phytophthora* root rot, and anthracnose stem and pod blights caused by *Colletotrichum lupini* causes several million dollars of losses in Australia, the largest producer of NLL globally (Sinden et al., 2004; Murray and Brennan, 2012). Considering the demonstrated antifungal activity of some seed storage proteins from several legume species, it was of interest to determine if seed storage proteins such as β -conglutins from NLL may also have roles in protection against fungal pathogens. Therefore the antifungal activity of NLL β -conglutins was examined using both *in vitro* and *in planta* assays for protection against fungal

¹<http://www.lupinexpress.org/>

and oomycete pathogen growth known to induce necrotic host tissue damage. Furthermore, insight into the potential inhibitory mechanisms by which these proteins act against pathogens was obtained through an assessment of their subcellular localization and impact on plant oxidative processes.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Plant experiments were conducted with *Nicotiana benthamiana* accession “lab” (Bally et al., 2015) in temperature controlled growth rooms as described by Petrie et al. (2010). Plants were grown under a 16-h light/8-h dark cycle at 22°C.

Fungal Isolates

Details of fungal isolates are listed in **Table 1** and were maintained as pure cultures with *Rhizoctonia solani*, *Alternaria brassicicola*, *F. oxysporum*, *Phytophthora nicotianae* and *C. lupini* isolates grown on 1/2 strength Potato Dextrose Agar (PDA), and *S. sclerotiorum* on 1/8 strength PDA. Spores, mycelia or sclerotia were inoculated on PDA plates, which were placed at room temperature in the dark until plates were fully covered by the pathogen. Mycelial plugs from these plates were used for subsequent experiments.

Construction of Expression Plasmids

β 1- and β 6-conglutins were overexpressed using the pET28b construct (Novogen)² that contains an N-terminal polyhistidine (6xHis) tag. pUC57 vectors carrying synthesized β 1, β 2, β 3, β 4, or β 6 conglutin sequences based on Genbank HQ670409 (β 1), HQ670410 (β 2), HQ670411 (β 3), HQ670412 (β 4) and HQ670414 (β 6) sequences but altered for optimum bacterial codon usage with *NcoI/XhoI* restriction enzyme linkers were synthesized and constructed by GenScript³ (**Supplementary Figure S1**). The bacterial expression vectors for β -conglutins were obtained via *NcoI/XhoI* digestion of respective pUC57- β -conglutin constructs followed by ligation of the β -conglutin fragments into the pET28b vector.

Overexpression and Purification of NLL β -Conglutin Proteins

All β -conglutin proteins were expressed in RosettaTM 2(DE3) pLysS SinglesTM Competent Cells (Novogen). Protein expression was performed using an auto-induction method (Studier, 2005). Briefly, a single clone containing the expression construct was isolated and grown for 20 h in LB plus kanamycin at 50 μ g/mL at 37°C and continuous shaking (200 rpm). The culture was diluted 1:150 in ZYM-5052 medium and grown for a further 5 h until the cell density reached an OD₆₀₀ of 0.7. The cells were then induced to overexpress the proteins by adjusting the temperature to 19°C for another 20 h. Cells were collected by centrifugation at 5000 \times g at 4°C. The bacterial cell pellet was rinsed two times with phosphate buffered saline (PBS), pH 7.5, removing

the supernatant, then flash frozen in liquid nitrogen and stored at -80°C until further use.

Purification of Recombinant β 1- and β 6-Conglutin Proteins

Protein purification from bacterial pellets was performed following the manufacturers' recommendations for His-tagged proteins (Qiagen)⁴. Briefly, the steps consisted of lysing cells followed by nickel affinity chromatography using Ni-NTA spin columns, and histidine (6xHis) tags at the N-terminal part of the β -conglutin proteins. After elution of 6xHis-tagged proteins from the column with an increasing imidazole concentration gradient (10–300 mM), 2.5 mL fractions were collected. Fractions containing protein were analyzed using SDS-PAGE and fractions showing a single band corresponding to the expected molecular weight were pooled, and dialyzed five times against Tris-HCl 100 mM, pH 7.5, 150 mM NaCl to eliminate the imidazole reagent. The protein was concentrated using a 30 kDa Amicon centrifuge filter (Millipore)⁵. The aliquots were flash-frozen in liquid nitrogen and kept at -80°C until further use. Protein purities were >95% as determined by densitometry analysis of the SDS-PAGE gel image. An aliquot of each protein was used to measure their concentration using Bradford assays (BioRad, Hercules, CA, USA) using bovine serum albumin (BSA) as a standard. The β -conglutins purifications yields ranged between 10–15 mg/mL.

β -Conglutin Antibody Production

The peptide sequence Nt – VDEGEGNYELVGIR – Ct, was chosen as this region was 100% homologous among all NLL β -conglutins and did not share any significant homology to other known lupin sequences. This peptide was generated by Agrisera⁶ and was used to immunize rabbits and to produce polyclonal antiserum (Agrisera). The rabbit immune serum was affinity-purified against the same synthetic peptide.

SDS-PAGE and Immunoblotting

SDS-PAGE and Immunoblotting were performed as previously described (Foley et al., 2015).

In vitro Assays for Fungal Growth Inhibition

A disk diffusion method was performed on 90 mm petri dishes containing PDA to test the sensitivity of different fungi strains toward the β -conglutin proteins. Fungal isolates were initially grown on PDA plates as described previously at 21°C until mycelial growth had developed. A mycelial plug was then taken from the growing edge of the colony and placed in the center of a new full-strength PDA plate in which sterile blank paper disks (12.7 mm diameter) containing 800 μ g of purified β -conglutin protein (dissolved in BSA buffer) or buffer only control were placed 30 mm away. For *in vitro* assay of *C. lupini*, 1 mg

² www.novogen.com

³ www.genscript.com

⁴ www.qiagen.com

⁵ http://www.emdmillipore.com

⁶ http://www.agrisera.com/

TABLE 1 | Details of fungal isolates used to assess β -conglutin antifungal activity.

Fungal pathogen	Isolate	Host(s)	Isolated off	Reference
<i>Rhizoctonia solani</i> AG8-1	WAC10335	Lupin, cereals, Brassicas	Lupin	Perl-Treves et al., 2004; Hane et al., 2014
<i>Rhizoctonia solani</i> AG2-1	WAC9767	Lupin, Brassicas	Lupin	Perl-Treves et al., 2004
<i>F. oxysporum</i> f. sp. <i>medicaginis</i>	Fom-5190a	<i>Medicago</i> species	Alfalfa	Williams et al., 2016
<i>F. oxysporum</i> f. sp. <i>conglutinans</i>	Fo5176	Brassicas	Cabbage	Thatcher et al., 2012
<i>Alternaria brassicicola</i>	UQ4273	Brassicas	Cabbage	Schenk et al., 2000
<i>Sclerotinia sclerotiorum</i>	UQ3833	Dicots, broad host range	Canola	Supplied by Kemal Kazan, CSIRO
<i>Colletotrichum lupini</i>	WAC8672	Lupins	NLL	Supplied by Julie McClements, DAFWA
<i>Colletotrichum lupini</i>	WAC10444	Lupins	NLL	Supplied by Julie McClements, DAFWA
<i>Phytophthora nicotianae</i>	PAB12.23	<i>Solanaceae</i> family	Tobacco	Supplied by Giles Hardy, Murdoch University.

of purified β 6-conglutin protein was used. The plates were incubated in the dark at 21°C and the zone of fungal inhibition around the disks recorded over 30 days. Assays were performed in triplicate.

Antifungal activity of β -conglutin proteins were expressed as the IC₅₀ (μ M) values for the fungi tested in Table 1. The mycelial growth inhibition assays were used to determine the concentration required for 50% growth inhibition (IC₅₀), using a β -conglutin protein concentration range from 5 to 125 μ M (5, 10, 15, 20, 25, 35, 50, 75, and 125 μ M) for each β -conglutin. Results were expressed as mean \pm standard deviation (SD). To determine the statistical significant differences of the β -conglutins antifungal activity on the growth of these fungal pathogens, the data was analyzed using statistical package SPSS 15.0 (SPSS, Inc., Chicago, IL, USA). Significant differences between the mean values of each cohort were determined using Tukey Kramer HSD test ($p < 0.05$). To characterize the antifungal activity of β -conglutins further, a subsequent experiment with the necrotrophic fungal pathogen of NLL causing anthracnose disease (*C. lupini* isolates WAC8672 and WAC10444) was conducted, where the antifungal activity of β 6-conglutin was determined by placing a mycelial plug in the center of a PDA plate and radial outgrowth determined over 14 days. For each time point significant differences were determined using a student's *t*-test using the JMP software v7.0 (SAS Institute).

Agroinfiltration of *Nicotiana benthamiana* Leaves

Four-week-old *N. benthamiana* plants grown at 22–24°C in culture rooms were used for *Agrobacterium tumefaciens*-mediated transient expression as described previously (Sparkes et al., 2006). The β 1- and β 6-conglutin coding sequence were cloned into the vector pMDC83 to generate C-terminal GFP fusion proteins and transformed into *A. tumefaciens* AGL1. Transformed *A. tumefaciens* AGL1 were cultured at 28°C until stationary phase (~24 h), washed and resuspended in infiltration medium (50 mM MES, 0.5% (w/v) glucose, 100 μ M acetosyringone (Sigma-Aldrich⁷ pH 5.6). The bacterial suspension was inoculated using a 1-mL syringe without a needle by gentle pressure through a <1 mm hole punched on the lower epidermal surface of the upper leaves of *N. benthamiana* plants.

⁷ www.sigmaaldrich.com

Following infiltration, plants were incubated under normal growth conditions at 22–24°C. This protocol was used for *in vivo* fungal growth inhibition assays, oxyblot assays, and subcellular localization studies.

Trypan Blue Staining for Fungal Hyphae and Dead Plant Cells

Forty hours after *Agrobacterium* infiltration of β 1- or β 6-conglutin protein constructs into *N. benthamiana* plants, agar plugs containing hyphae of *S. sclerotiorum* or *P. nicotianae* were placed on the infiltrated leaf areas (control and β 1- or β 6-conglutin overexpression) and plants were incubated in a growth chamber at 18°C for 3 days under a 16 h long-day light regime. Leaves were assessed for visible necrotic disease progression, then detached and further visualized after lactophenol trypan blue staining based on Keogh et al. (1980). Briefly, mature fourth leaves of *N. benthamiana* containing visible infection sites (tissue necrosis) were cleared with acetic acid: ethanol (1:1 v/v) then stained for 1–2 h in lactophenol (10 mL phenol; 10 mL lactic acid, 10 mL water) with 0.05% (w/v) trypan blue at 60°C. Excess staining was removed with lactic acid: water (1:1 v/v) until leaves were clear. Leaves were examined by light microscopy on a Nikon N400-M light microscope (Nikon, Tokyo, Japan). Control areas were infiltrated with GFP-only vectors. The experiment was repeated three times. In each experiment, leaves from 4–6 plants were analyzed for each treatment.

Subcellular Localization of β 1- and β 6-Conglutin in Plant Cells

Fusion proteins were expressed in 3-week-old *N. benthamiana* via *Agrobacterium* infiltration of leaves as previously described (Sparkes et al., 2006). Leaves were excised 3 days following infiltration and mounted with water under a 0.17 mm coverslip and imaged using a Nikon A1Si confocal microscope (Nikon Plan Apo VC 60x NA1.2 water-immersion objective). For GFP imaging, the 488 nm laser line and a 521/50 nm band pass filter was utilized, while a 561 nm laser line and 595/50 nm filter was used for RFP imaging.

Images were analyzed using ImageJ software (Schneider et al., 2012). Images were converted to 8-bit grayscale and the intensity correlation analysis (ICA) method was used to determine the levels of colocalization and by using the JACoP plugin according

to (Bolte and Cordelières, 2006). As a control, empty vector was used to transform leaf cells expressing 35S::GFP alone as described previously by Thatcher et al. (2007).

Oxyblot Assays

Proteins were extracted from *N. benthamiana* leaves infiltrated with GFP, β 1-GFP or β 6-GFP fusion expression constructs following either control or *S. sclerotiorum* or *P. nicotianae* treatments [extraction buffer: 25 mM Tris-HCl, pH 7.0, 0.05% Triton X-100, 1 mM dithiothreitol (DTT), and protease inhibitors (Roche, Basel, Switzerland)]. 25 μ g of total proteins were loaded onto 12% polyacrylamide gels for protein separation. Proteins separated by SDS-PAGE were electrotransferred to PVDF membranes. The OxyBlotTM Protein Oxidation Detection Kit (EMD Millipore) was used according to the manufacturer's instructions for immunoblot detection of carbonyl groups introduced into proteins by reaction with reactive oxygen species (ROS).

RESULTS

In vitro Inhibition of Fungal Growth by β -Conglutin Recombinant Proteins

To assess the potential for antifungal activity in NLL seed storage proteins, we focussed on the most abundant seed storage proteins in the NLL grain, the β -conglutins (Foley et al., 2011). The 6xHis-tag recombinant β -conglutin proteins were expressed in *E. coli* and purified using nickel affinity chromatography. To confirm the identity of purified β -conglutins, SDS-PAGE analysis of the purified proteins was performed, which indicated a single protein band of approximately 65 kDa, which is the predicted size of β -conglutin (Supplementary Figure S2A). This was followed by immunoblotting using an anti- β -conglutin antibody which confirmed the identity of the recombinant proteins as β -conglutin (Supplementary Figure S2B). We were successful in expressing and purifying the β 1, β 2, β 3, β 4 and β 6 recombinant proteins but not β 5 and β 7.

Subsequently, we examined the effect of the purified β -conglutin proteins on the growth rate of a range of phytopathogenic necrotrophic fungi using *in vitro* bioassays. The fungal pathogens selected included the legume pathogen *F. oxysporum* forma specialis (f. sp.) *medicaginis* (Fom-5190a, a root pathogen), the broad host range pathogens *R. solani* AG8-1 (isolated from lupin) and *S. sclerotiorum* (isolated from canola) and the brassica-specific pathogens *F. oxysporum* f. sp. *conglutinans* (Fo-5176), *R. solani* AG2-1 and *A. brassicicola* (Brassicaceae hosts; details of these pathogens are listed in Table 1). The IC₅₀ values (μ M) for each of these fungal isolates was determined (Table 2). Overall the β 1 and β 6 conglutins showed significantly stronger mycelium growth inhibition when compared to β 2-, β 3-, and β 4-conglutin proteins for both *R. solani* isolates and *A. brassicicola* by Tukey-Kramer honestly significant difference (HSD) test ($P < 0.05$). The β 1-conglutin showed a significantly stronger growth inhibition to *S. sclerotiorum* and the two *F. oxysporum* isolates, compared to β 2-, β 3-, and β 4-conglutin, where β 6-conglutin was not

significantly different from β 1 (Table 2). Overall β 1-, and β 6-conglutin showed the strongest mycelial growth inhibition to the various pathogens tested, but interestingly, a sequence alignment of the seven β -conglutin proteins showed β 6 exhibits the highest sequence identity to other β -conglutin protein isoforms (78–98%) while β 1 had an amino acid sequence with the lowest identity (77–81%; Supplementary Figure S3). Control treatment of filter disks with BSA buffer showed no fungal growth inhibition against any of the isolates tested.

Based on the β -conglutin protein alignments and IC₅₀ data, we decided to focus on β 6-conglutin as a representative member of the β -conglutin family. Further characterization of the *in vitro* anti-fungal properties of β 6-conglutin was performed in a detailed time course experiment against two isolates (WAC8672 and WAC10444) of a major fungal pathogen of lupins, *C. lupini* which causes anthracnose disease (Fischer et al., 2015). Radial outgrowth of *C. lupini* mycelium on PDA plates toward Whatman filter disks containing control protein (BSA), β 6-conglutin or no protein was recorded. Radial growth inhibition was only observed toward filter disks containing β 6-conglutin protein and this occurred from as early as 6 days post-inoculation with the mycelial plug for isolate WAC10444 and 8 days post-inoculation for WAC8672 (Figure 1). Combined with the IC₅₀ data in Table 2, these results indicate recombinant NLL β -conglutins exhibit antifungal activity *in vitro* against both leaf and root-infecting pathogens of legumes and non-legume hosts.

β 1- and β 6-Conglutins Exhibit *in planta* Anti-fungal and Oomycete Activity

To examine the effect of NLL β 6-conglutin *in planta*, we selected the *N. benthamiana* infiltration system as a model for assessing the functionality of proteins against various phytopathogens (Ma et al., 2012). This involved *Agrobacterium*-mediated infiltration into leaves of *N. benthamiana* plants followed by assessment of antifungal activity in disease assays. The broad host range leaf pathogen *S. sclerotiorum* was chosen which is readily amenable to *N. benthamiana* leaf disease assays and secretes the non-host selective toxin oxalic acid to induce disease symptom development (Kim et al., 2008; Williams et al., 2011). In addition, β 6-conglutin showed strong inhibition of this pathogen's growth as compared to other fungal pathogens tested in our *in vitro* assays (Table 2). The *Agrobacterium*-infiltration system was used to transiently express β 6-GFP or a GFP-only control. The GFP control was infiltrated into one half of the leaf with β 6-GFP infiltrated into the other leaf half. Forty-eight hours after infiltration, leaves were inoculated with *S. sclerotiorum* and progression of lesions was observed over 72 h (Figure 2). In the GFP-only control, necrosis was apparent within 24 h of *S. sclerotiorum* inoculation with the size of the necrotic lesions progressing rapidly to engulf half of the leaf by 72 h. In stark contrast, there was only limited necrotic damage in the leaves expressing the β 6-GFP protein. As β 1-conglutin demonstrated strong antifungal activity *in vitro* yet exhibits the least amino acid identity amongst the NLL β -conglutins, we also assayed β 1-GFP in the above experiments. As with β 6-GFP, β 1-GFP infiltrated

TABLE 2 | Antifungal activity of the NLL recombinant purified β 1- to β 4-conglutins and β 6-conglutins.

Fungi species	Isolate	IC50 values (μ M)				
		β 1	β 2	β 3	β 4	β 6
<i>Rhizoctonia solani</i> AG8-1	WAC10335	18.5 \pm 2.5 ^c	43.0 \pm 4.0 ^{ab}	36.8 \pm 3.7 ^b	52.5 \pm 5.5 ^a	20.5 \pm 2.0 ^c
<i>Rhizoctonia solani</i> AG2-1	WAC9767	17.4 \pm 2.0 ^c	45.2 \pm 3.5 ^a	31.5 \pm 3.8 ^b	46.2 \pm 4.1 ^a	23.2 \pm 1.7 ^c
<i>Sclerotinia sclerotiorum</i>	UQ3833	14.0 \pm 1.3 ^c	27.0 \pm 4.2 ^a	23.6 \pm 2.0 ^{ab}	26.3 \pm 3.2 ^{ab}	19.3 \pm 2.4 ^{bc}
<i>Alternaria brassicicola</i>	UQ4273	16.0 \pm 1.7 ^b	34.0 \pm 4.6 ^a	24.0 \pm 3.3 ^b	37.8 \pm 3.0 ^a	21.6 \pm 2.2 ^b
<i>F. oxysporum</i> f. sp. <i>conglutinans</i>	Fo5176	18.7 \pm 1.5 ^c	41.3 \pm 5.8 ^a	29.7 \pm 2.5 ^b	40.1 \pm 3.8 ^a	25.5 \pm 3.5 ^{bc}
<i>F. oxysporum</i> f. sp. <i>medicaginis</i>	Fom-5190a	20.3 \pm 2.7 ^c	39.0 \pm 5.2 ^{ab}	32.0 \pm 4.2 ^b	44.3 \pm 5.0 ^a	29.7 \pm 4.0 ^{bc}

Protein concentrations (μ M) required for IC50 were determined from the dose-response curves (percentage of growth inhibition versus protein concentration). The results are expressed as mean \pm standard deviation (SD) of three biological replicates. Statistically significant differences were calculated using a Tukey-Kramer honestly significant difference (HSD) test ($P < 0.05$). Different letters for each of the different pathogens indicate significant difference in IC50 value of the β -conglutin.

leaves also exhibited limited necrotic damage after *S. sclerotiorum* inoculation (Supplementary Figure S4).

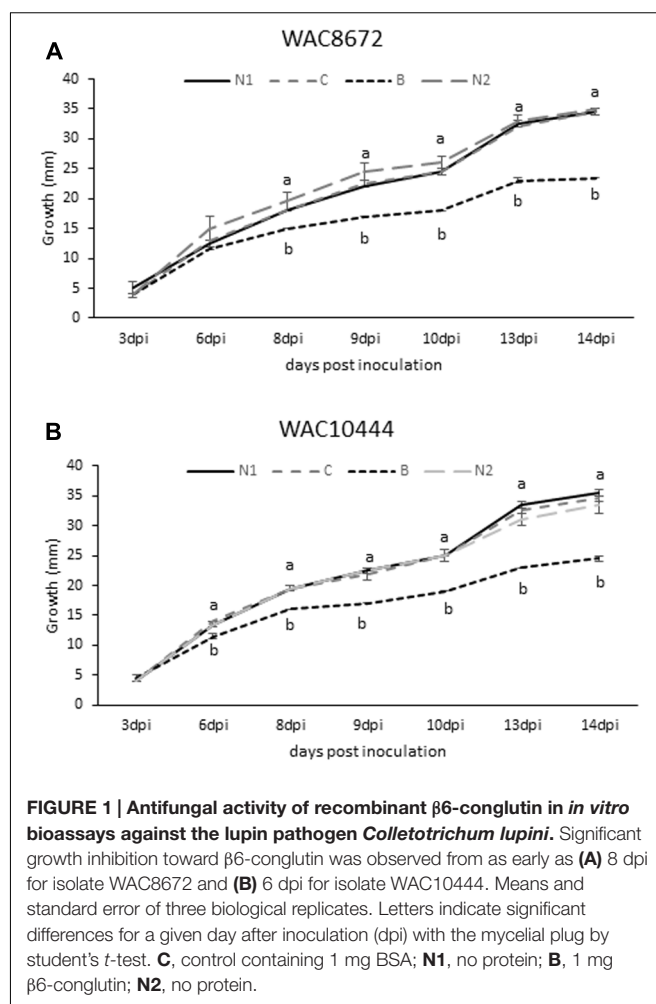
The soil-borne oomycete pathogen of *N. benthamiana*, *P. nicotianae*, was also assayed. *P. nicotianae* is a hemibiotrophic pathogen that causes root rot, leaf necrosis and stem lesions (Liu et al., 2016). As with the *S. sclerotiorum* assays, both β 6- and β 1-conglutin strongly inhibited lesion development by *P. nicotianae* in our *N. benthamiana* *Agrobacterium*-infiltration disease assays (Figure 2 and Supplementary Figure S4). At 72 h post *S. sclerotiorum* or *P. nicotianae* inoculation, β 6-conglutin infiltrated leaf zones exhibited a 71.7 and 85.7% reduction in lesion size, respectively, relative to control treated leaf zones. Similar reductions were recorded for β 1-conglutin infiltrated leaf zones (94.2 and 90.3%).

β 1- and β 6-Conglutin Reduce Pathogen Growth and Pathogen Induced Cell Death *In planta*

The striking inhibition of *S. sclerotiorum* and *P. nicotianae* induced lesions on *N. benthamiana* leaves expressing β 1- or β 6-conglutin suggests these pathogens are unable to grow or their growth is severely impaired by these β -conglutins. To examine fungal/oomycete growth we challenged *N. benthamiana* β -conglutin expressing leaves with pathogen and allowed 72 h for disease symptom development in controls, then assessed the leaves and mycelial growth microscopically after staining with trypan blue, which stains dead plant cells (van Wees, 2008). In leaves expressing β 6-GFP or β 1-GFP we observed strong inhibition of the *S. sclerotiorum* and *P. nicotianae* mycelial growth compared to controls (Figure 3 and Supplementary Figure S5). There was evidence of aggregated cells with short pseudohyphae, particularly at the site of inoculation, however, these were few and sparse in comparison to controls. Combined, our results suggest β 1- and β 6-conglutins from NLL inhibit hyphal growth of a range of phytopathogenic oomycete and fungi, both *in vitro* and *in vivo*.

Subcellular Localization of β -Conglutin in *N. benthamiana* Leaves

The effect of β -conglutins on the growth of fungal and oomycete pathogens tested led to the hypothesis that β 1- and β 6-conglutins



might localize at the cell surface, at sites closest to initial pathogen attack. Therefore, the subcellular localization of the β -conglutin proteins using *Agrobacterium*-mediated transient expression of β -GFP constructs in *N. benthamiana* leaves was examined. Two-to-three days following *Agrobacterium* infiltration, the localisation of β -GFP in *N. benthamiana* leaf epidermal cells was examined by confocal microscopy. Both

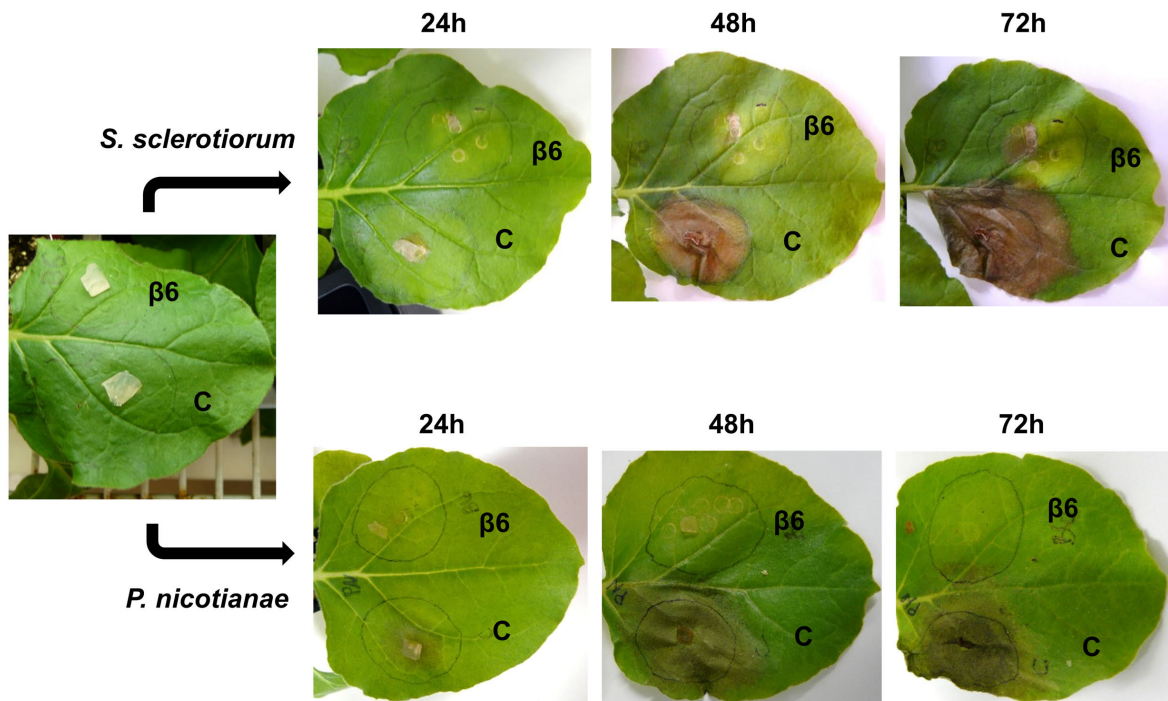


FIGURE 2 | Recombinant β 6-conglutin exhibits *in planta* anti-fungal and oomycete activity. Shown are representative images of *Agrobacterium* infiltrated *N. benthamiana* leaves expressing recombinant β 6-conglutin proteins and subsequently inoculated with either *S. sclerotiorum* or *P. nicotianae*. The experiment was repeated three times with similar results. C, control agroinfiltration of the leaf area with *Agrobacterium* expressing GFP only; β 6, agroinfiltration of the leaf area with *Agrobacterium* expressing GFP tagged β 6-conglutin.

β 6-GFP and β 1-GFP were expressed in punctate structures close to the cell surface (plasma membrane; **Figure 4**, **Supplementary Figure S6**). To determine the nature of the β 6-GFP- and β 1-GFP-positive structures, which resembled the pattern described for plasmodesmata (Lee and Lu, 2011), we co-expressed the constructs with the known plasmodesmata marker plasmodesmata-located protein1 or AtPDL1-mCherry (Thomas et al., 2008). β 6-GFP and β 1-GFP partially overlapped the expression pattern of AtPDL1-mCherry ($R = 0.0725$ – 0.755), indicating that β 6-GFP and β 1-GFP were partially located at plasmodesmata (**Figure 5**, and **Supplementary Figure S6**).

Protein Oxidation Levels in *N. benthamiana* Leaves Expressing the β 1- and β 6-Conglutin Proteins Following Infection with *S. sclerotiorum* and *P. nicotianae*

One of the mechanisms employed by *S. sclerotiorum* and *P. nicotianae* during infection of a compatible host is to initially suppress the plant second-phase oxidative burst that occurs 3–6 h after pathogen contact (Levine et al., 1994; Cessna et al., 2000), thereby compromising the capacity of the plant to activate downstream defense pathways (Doke, 1985; Levine et al., 1994; Williams et al., 2011). Therefore the effect of β -conglutin protein on the capacity of *S. sclerotiorum* and *P. nicotianae* to suppress

the plant oxidative burst was examined. Leaves were infiltrated and allowed to transiently express GFP or β 6-GFP conglutin for 48 h before being infected with *S. sclerotiorum* or *P. nicotianae*. As soon as hyphae and lesions became visible (within 24 h) leaves were collected for analysis (as shown in **Figure 6A**). We estimated production of ROS and oxidative burst capacity by examining the level of protein carbonylation in infected compared to mock-inoculated leaves using an OxyBlot Protein Oxidation Detection and immunoassay (Rinalducci et al., 2008). Protein oxidation is one of the covalent modification of proteins induced by ROS such as H_2O_2 or other products of oxidative stress, and carbonylation is one of the most commonly occurring oxidative modifications of proteins, which may be responsible for the alteration in protein activity, for example, signaling (Oracz et al., 2007). Carbonylated proteins have been identified in many plant species at different stage of growth and development (Barba-Espín et al., 2011; Morscher et al., 2015).

Basal levels of protein oxidation, as generated through normal metabolic activity (Alscher et al., 1997; Rinalducci et al., 2008) were observed in the mock-inoculated control leaves expressing GFP-only, as well as in mock-inoculated leaves expressing β 6-GFP (**Figures 6B–C**). Following inoculation with *S. sclerotiorum* or *P. nicotianae* protein oxidation remained at similar levels in the GFP-only control leaves (**Figures 6B–C**). In contrast, we observed a marked increase in the levels of protein oxidation in leaves expressing the β 6-GFP following infection with *S. sclerotiorum* or *P. nicotianae* when compared

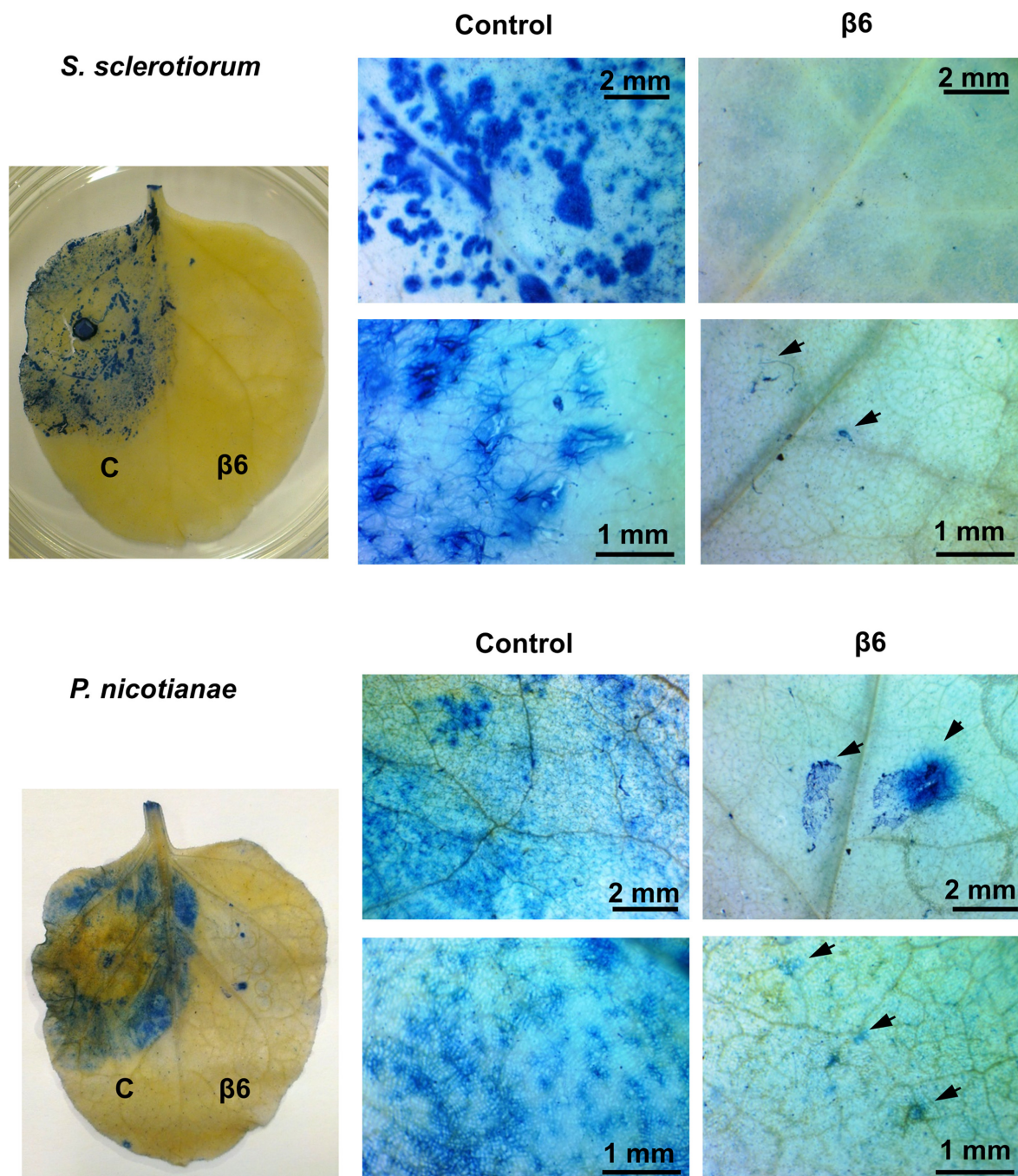


FIGURE 3 | Recombinant $\beta 6$ -conglutin reduces pathogen growth and pathogen induced cell death *in planta*. Shown are representative images of *Agrobacterium* infiltrated *N. benthamiana* leaves expressing recombinant $\beta 6$ -conglutin proteins and subsequently inoculated either *S. sclerotiorum* or *P. nicotianae*. Trypan blue staining was performed to visualize hyphal growth and cell death. Arrows point to hyphae and hyphal damage.

to the respective mock-inoculated $\beta 6$ -GFP or the infected GFP-only leaves. The inoculated leaves expressing $\beta 6$ -GFP were nevertheless healthy, as expected (Figure 6A). This suggests that the over-expression of β -conglutin proteins effectively circumvents the initial suppression of the plant oxidative burst by *S. sclerotiorum* or *P. nicotianae*.

DISCUSSION

β -conglutins are the most abundant seed storage proteins in NLL (Foley et al., 2011) and while in other plant species these vicilin-like proteins may have roles in plant defense, the functional roles of β -conglutins in this aspect remain

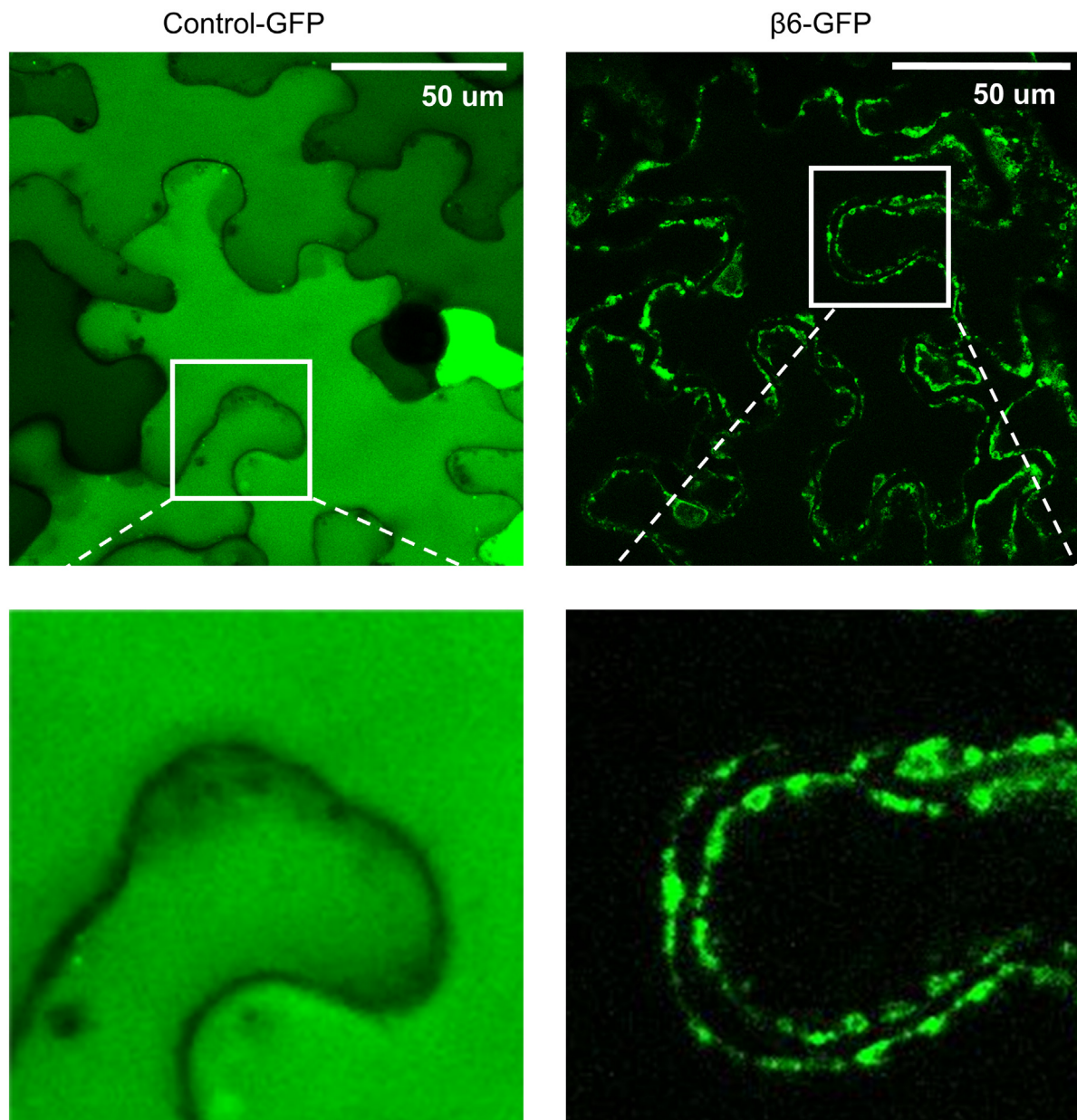


FIGURE 4 | β 6-conglutin is localized to the cell surface. Confocal images of tobacco epidermis cell expressing GFP alone or β 6-GFP shows GFP alone homogeneously expressed throughout the cytoplasm while β 6-conglutin localizes to the plasma membrane through the whole cell, with no expression in cytosol or intracellular organelles. Insert: β 6-GFP shows punctate labeling at the cell surface.

largely unknown (Khuri et al., 2001; Dunwell et al., 2004). In this study we identified two NLL β -conglutin proteins that strongly inhibited the growth of a range of necrotrophic fungal or oomycete pathogens, both *in vitro* and *in vivo* when transiently expressed in *N. benthamiana* leaves. Reduced *in planta* fungal growth was associated with a significant reduction in pathogen-induced host cell death and interestingly the NLL β -conglutins examined were localized near the plant cell surface. These results provide the first demonstration for any NLL β -conglutin in protection against pathogen attack, and add to

the growing list of vicilin-like proteins that accumulate during seed development and have roles in plant defense (Gomes et al., 1998; Marcus et al., 1999; Rietz et al., 2012; Monteiro et al., 2015).

Vicilin-like proteins are members of the cupin superfamily which is extremely diverse, encompassing 18 different functional classes including the vicilins and similar germin-like seed storage proteins, as well as single-barrel isomerases, epimerases, and auxin-binding proteins (Dunwell et al., 2001). Given the varying antifungal potency of vicilin and vicilin-like proteins from

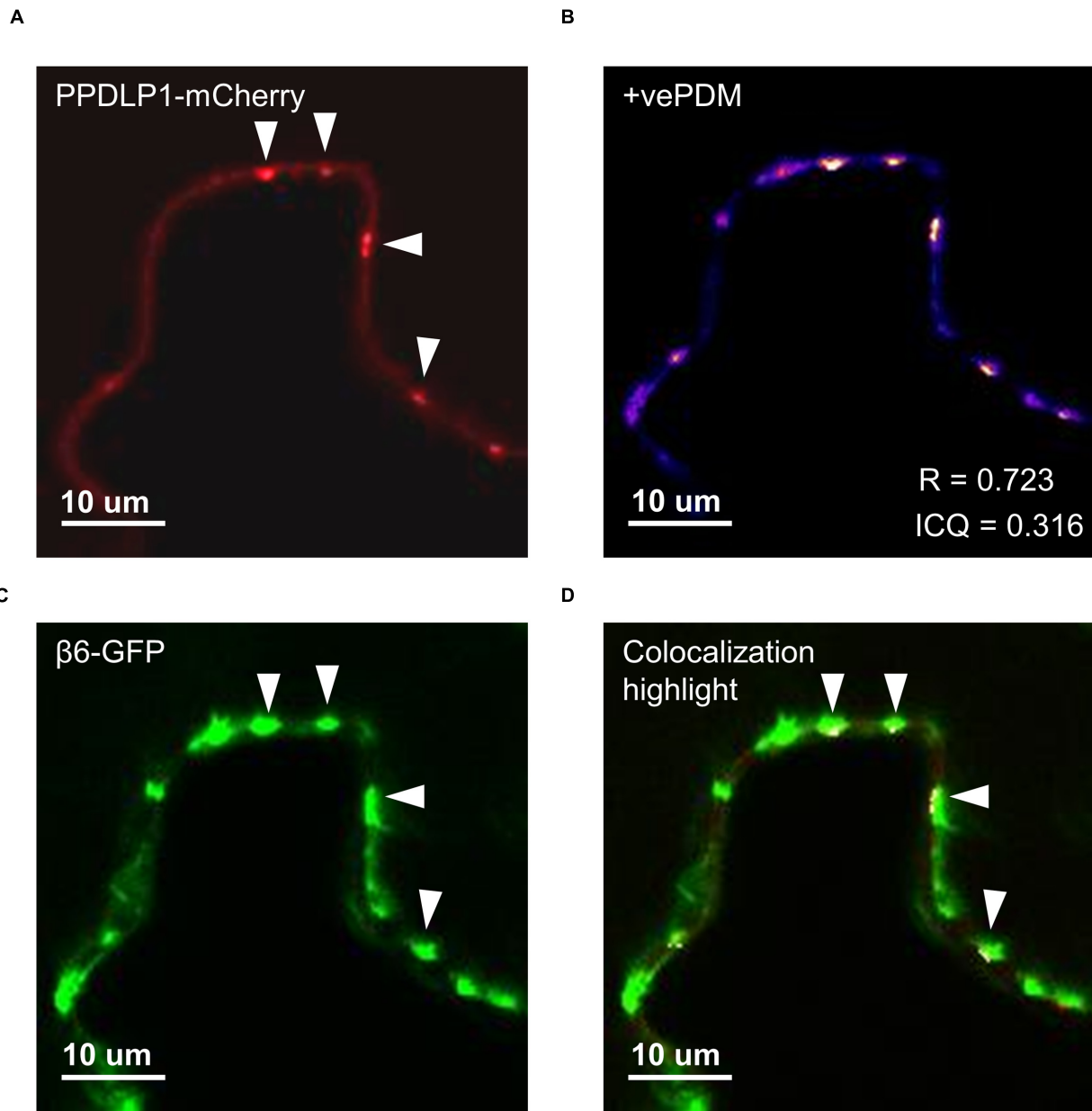


FIGURE 5 | β 6-conglutinin localizes to the plasmodesmata. Single-slice confocal images of co-expression GFP- β 6 with the plasmodesmata marker PDL1-mCherry after transient expression in *N. benthamiana*; **(A)** PDL1-mCherry, **(B)** β 6-GFP, **(C)** Image showing pixel pairs that have a positive PDM value equal to the value (intensity of A – mean A intensity) * (intensity of B – mean B intensity) as described in Li et al. (2004), **(D)** merge of **(A,B)** with highlighted co-localized pixels. ICQ, Intensity correlation quotient; R, Mandel's overlap coefficient. 60 \times immersion objective.

various plant species (Gomes et al., 1998), to determine and compare and contrast the ability of NLL β -conglutins to inhibit fungal growth we assayed each of the five synthesizable NLL β -conglutins against a range of necrotrophic pathogens. Of the five NLL β -conglutins, β 1 and β 6 exhibited the strongest activity *in vitro*. Sequence comparisons among the β -conglutins does not reveal any motif common between β 1 and β 6 but not in the other tested β -conglutins (**Supplementary Figure S3**) so at this stage we are unable to hypothesize why β 1 and β 6 have stronger fungal

inhibition activity than the other β -conglutins. Furthermore, we demonstrated β -conglutinin antifungal activity *in planta* against *S. sclerotiorum* as well as against a hemibiotrophic oomycete pathogen, *P. nicotianae*. Necrotrophic fungal pathogens actively kill host tissue, while hemibiotrophic pathogens switch to this attack mode during later stages of their infection cycle (Glazebrook, 2005).

Although plant defense responses against pathogen attack are the result of various integrated preformed and induced

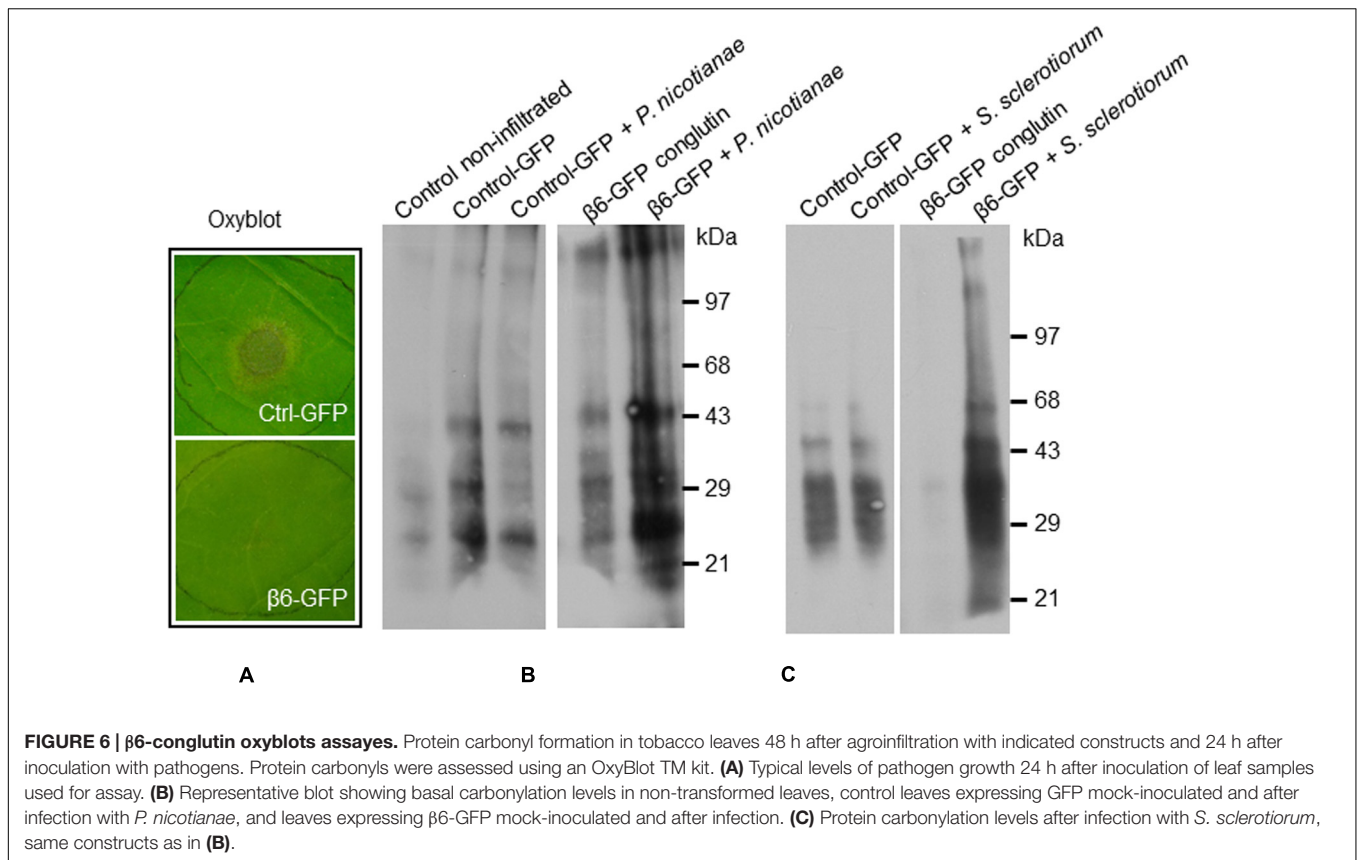


FIGURE 6 | β 6-conglutin oxyblots assays. Protein carbonyl formation in tobacco leaves 48 h after agroinfiltration with indicated constructs and 24 h after inoculation with pathogens. Protein carbonyls were assessed using an OxyBlot TM kit. **(A)** Typical levels of pathogen growth 24 h after inoculation of leaf samples used for assay. **(B)** Representative blot showing basal carbonylation levels in non-transformed leaves, control leaves expressing GFP mock-inoculated and after infection with *P. nicotianae*, and leaves expressing β 6-GFP mock-inoculated and after infection. **(C)** Protein carbonylation levels after infection with *S. sclerotiorum*, same constructs as in **(B)**.

mechanisms, one of the most prominent is the hypersensitivity response (HR) resulting from the generation of host ROS (Mittler, 2002). While the HR response is a type of programmed cell death that can limit the growth of biotrophic pathogens, it is favorable to necrotrophic pathogens that thrive off the dead host cells (Glazebrook, 2005; Laluk and Mengiste, 2010). Both *S. sclerotiorum* and *P. nicotianae* are capable of inciting necrotic lesions on a broad range of host plants (Agrios, 2005; Gallup et al., 2006) where *S. sclerotiorum* produces the major pathogenicity factor oxalic acid (Cessna et al., 2000), the primary determinant contributing to its pathogenic success (Kim et al., 2011). In compatible interactions, oxalic acid initially dampens the plant oxidative burst (Williams et al., 2011). However, once the pathogen is established, oxalic acid induces apoptotic-like programmed cell death in plant hosts, triggered by the generation of ROS at detrimental levels (Kim et al., 2008). We found *in planta* expression of NLL β 1- and β 6-conglutins effectively impaired host cell death induced by both *S. sclerotiorum* and *P. nicotianae*, evident within 24 h of pathogen challenge and lasting over the 72 h assayed. *In planta* expression of NLL β 1 and β 6-conglutins also increased levels of pathogen (*S. sclerotiorum*, *P. nicotianae*) induced protein oxidation whilst maintaining leaf health, suggesting overexpression of these two β -conglutins inhibits pathogen induced suppression of the early phase plant oxidative burst.

To dissect how NLL β -conglutins inhibit pathogen growth and host cell death *in planta*, we utilized GFP-tagged versions

of these proteins to visualize their sub-cellular localisation. The vacuolar localisation of vicilin-like proteins (to supply amino acids during seed germination and seedling growth) has been extensively reported, however, almost no studies have been conducted for these protein classes in organs other than seeds (Overvoorde et al., 1997). Germin-like proteins from peanut localize to both the cytoplasm and the cell surface (cell membrane or cell wall) when transiently expressed within onion epidermal cells (Wang et al., 2013). Here we observed both the NLL β 1- and β 6-conglutin proteins localizing to the cell surface in distinct structures that included plasmodesmata when expressed in *N. benthamiana* leaf epidermal cells. Many studies have demonstrated that ROS are produced at the plant cell wall in a highly regulated manner (Wojtaszek, 1997; Sewelam et al., 2016), where they play key signaling roles in the control of physiological processes such as cellular growth and development (Gapper and Dolan, 2006; Kärkönen and Kuchitsu, 2015), as well as adaptation to environmental changes and pathogen attack (Wu et al., 1997; Sewelam et al., 2016). In plants one of the major contributors to ROS production during pathogen infection are the plasma membrane localized NADPH oxidases (Torres et al., 2006; reviewed in Schopfer and Liskay, 2006). It is possible that NLL β -conglutins facilitate/mediate the production of ROS directed to the oxidative burst (Bolwell et al., 1995; Bolwell and Wojtaszek, 1997), which is known to induce structural reinforcement of the cell wall through lignin crosslinking. This has been reported for

some cupins (germin and germin-like proteins) from wheat (Schweizer et al., 1999). Alternatively, ROS such as H_2O_2 could play direct antimicrobial roles or act as a signaling molecule in defense response pathways (reviewed in Shetty et al., 2008).

Structurally, the β -conglutin proteins are similar to germins, germin-like proteins, and vicilin-like glucose binding proteins, which are also glycoproteins characterized by a beta-barrel core structure that can be associated with the cell wall (Lane et al., 1992), the plasma membrane (Overvoorde et al., 1997; Kukavica et al., 2005) and/or plasmodesmata (Ham et al., 2012). The structure of β -conglutin is unique as it possesses two cupin domains forming a Rossmann fold reminiscent of enzymes that use molecular oxygen as a substrate (Jimenez-Lopez et al., 2015). Germins and germin-like proteins have been shown to play dual roles in seed germination and also in pathogen defense (Cândido et al., 2011). Identified from germinating wheat embryos, the wheat germin protein exhibits oxalate oxidase activity, catalyzing the conversion of oxalates (the conjugate base of oxalic acid) into CO_2 and H_2O_2 (Woo et al., 2000; Pan et al., 2007). Other enzymatic properties of germins or germin-like proteins include superoxide dismutase (SOD) activity, ADP glucose pyrophosphatase/phosphodiesterase activity or polyphenol oxidase (PPO) activity (reviewed in Barman and Banerjee, 2015). Over-expression of germin in several plant species can lead to increased resistance to fungal pathogens such as *S. sclerotiorum* (Donaldson et al., 2001; Dong et al., 2008; Walz et al., 2008), and the over-expression of a germin-like oxalate oxidase in rice or sunflowers lead to increased resistance, respectively, against *R. solani* (Molla et al., 2013) or both *R. solani* and *S. sclerotiorum* (Beracochea et al., 2015). Moreover, overexpression of the sunflower germin-like protein in Arabidopsis altered host redox and increased endogenous ROS levels (Beracochea et al., 2015). Germin-like proteins from *Brassica napus* have also been linked to the initiation of an oxidative burst that impedes pathogenesis of *S. sclerotiorum* (Rietz et al., 2012).

A role for β -conglutin in pathogen resistance has also been proposed based on cleavage and secretion of a β -conglutin peptide (BLAD) upon germination in *L. albus* (Monteiro et al., 2015). BLAD, a 20 kDa polypeptide, accumulates exclusively in the cotyledon between days 4 and 12 after the onset of germination. BLAD forms a 120 kDa oligomeric structure which exhibits lectin-like activity, catalytic activities of β -N-acetyl-D-glucosaminidase and chitin-binding activity, and provides effective antifungal activity against a range of plant pathogens (Monteiro et al., 2015). Whilst the results presented in our current study indicate the involvement of the NLL β -conglutin proteins in facilitating the production of ROS following pathogen infection *in planta*, it remains possible that some of the effects observed, particularly those obtained with the *in vitro* plate assays, may be partially linked to anti-fungal activities similar to those observed with the BLAD peptide. Protein exudates of germinating *L. albus* seeds showed fungal growth inhibition to five of six pathogens tested (Scarafoni et al., 2013). This protein exudate contains a range of different proteins including both

β - and γ -conglutins. Our research presented herein has shown that β -conglutins have antifungal activity and the β -conglutins from *L. albus* in the protein exudate could thus be a good candidate for contributing to the causal antifungal activity observed. It is therefore possible that lupins secrete β -conglutins during the vulnerable initial seedling germination stage as a means to protect itself from plant pathogens. As the BLAD peptide has been processed from β -conglutin, it remains to be determined if $\beta 1$ and $\beta 6$ would have altered antifungal properties if these proteins were also processed similar to that of BLAD.

The results presented herein suggest that NLL β -conglutins may be more versatile in their physiological roles than previously thought. While a clear causal connection cannot be given at present, our results show that several NLL β -conglutins inhibit fungal growth *in vitro* and that expression of at least two of these *in planta* enhances plant resistance to fungal/oomycete necrotrophic pathogens.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: JJ-L, SM, KD, and LT. Performed the experiments: JJ-L, SM, KD, LT, and LK. Analyzed the data: JJ-L, SM, KD, LT, LK, and KS. Contributed reagents/materials/analysis tools: RF, JJ-L, and KS. Wrote the paper: JJ-L, SM, KD, LT, RF, LK, and KS.

FUNDING

This work was supported by the European Research Program MARIE CURIE (FP7-PEOPLE-2011-IOF) for the grant ref. number PIOF-GA-2011-301550 to JJ-L and KS. JJ-L thanks the Spanish Ministry of Economy and Competitiveness for the grant ref. number RYC-2014-16536 (Ramon y Cajal Research Program). SM was supported by a CSIRO OCE Postdoctoral Fellowship.

ACKNOWLEDGMENTS

We thank Roger Shivas for the *F. oxysporum* f. sp. *conglutinans* strain Fo-5176, John Irwin for the *F. oxysporum* f. sp. *medicaginis* strain Fom-5190a (BRIP 5190a), Kemal Kazan (CSIRO) for the *A. brassicicola* (UQ4273) and *S. sclerotiorum* (UQ3833) isolates, Mark Sweetingham (DAFWA) for the *R. solani* isolates AG8-1 and AG2-1, and Prof. Giles Hardy (Murdoch University) for the *P. nicotianae* (PAB12.23). We thank the Department of Agriculture and Food of Western Australia for supplying the *C. lupini* isolates WAC8672 and WAC10444. We acknowledge the facilities, the scientific and technical assistance of the Australian Microscopy and Microanalysis Research Facility at the Centre for Microscopy, Characterisation and Analysis, The University of Western Australia, a facility funded by the University, State and Commonwealth Governments. We also thank Nicholas Pain for excellent technical assistance and TJ Higgins for helpful comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01856/full#supplementary-material>

FIGURE S1 | Sequences of synthetic β 1, β 2, β 3, β 4, and β 6 conglutins that were cloned into the expression vector, pET28b.

FIGURE S2 | Purification and confirmation of recombinant β 1- to β 4- and β 6-conglutins. (A) Purified proteins (10 μ g per sample) were separated by SDS-PAGE analyses to indicate a single protein band (6xHis-tag) of approximately 65 kDa at high purity level (>95%). **(B)** Immunoblot confirmation using anti- β -conglutinin protein antibody. The arrow indicates the correct sized band.

FIGURE S3 | Comparison of the NLL β -conglutinin protein sequences. Alignment of the seven β -conglutinin proteins identified in NLL, and comparison of the level of variability among them. Similarity percentages for each compared-pair of sequences are described in the table below. The lowest and the highest percentages of similarity are highlighted with gray color.

REFERENCES

- Agrios, G. N. (2005). "Plant diseases caused by fungi," in *Plant Pathology*, 5th Edn, ed. G. N. Agrios (San Diego, CA: Academic Press), 385–614.
- Alscher, R. G., Donahue, J. L., and Cramer, C. L. (1997). Reactive oxygen species and antioxidants: Relationships in green cells. *Physiol. Plant.* 100, 224–233. doi: 10.1111/j.1399-3054.1997.tb04778.x
- Argos, P., Narayana, S. V., and Nielsen, N. C. (1985). Structural similarity between legumin and vicilin storage proteins from legumes. *EMBO J.* 4, 1111–1117.
- Bally, J., Nakasugi, K., Jia, F., Hung, H., Ho, S. Y. W., Wong, M., et al. (2015). The extremophile *Nicotiana benthamiana* has traded viral defence for early vigour. *Nat. Plants* 1:15165. doi: 10.1038/nplants.2015.165
- Barba-Espin, G., Diaz-Vivancos, P., Job, D., Belghazi, M., Job, C., and Hernández, J. A. (2011). Understanding the role of H₂O₂ during pea seed germination: a combined proteomic and hormone profiling approach. *Plant Cell Environ.* 34, 1907–1919. doi: 10.1111/j.1365-3040.2011.02386.x
- Barman, A. R., and Banerjee, J. (2015). Versatility of germin-like proteins in their sequences, expressions, and functions. *Funct. Integr. Genomics* 15, 533–548. doi: 10.1007/s10142-015-0454-z
- Beracocha, V. C., Almasia, N. I., Peluffo, L., Nahirnak, V., Hopp, E. H., Paniego, N., et al. (2015). Sunflower germin-like protein HaGLP1 promotes ROS accumulation and enhances protection against fungal pathogens in transgenic *Arabidopsis thaliana*. *Plant Cell Rep.* 34, 1717–1733. doi: 10.1007/s00299-015-1819-4
- Berger, J. D., Clements, J. C., Nelson, M. N., Kamphuis, L. G., Singh, K. B., and Buirchell, B. (2013). The essential role of genetic resources in narrow-leaved lupin improvement. *Crop Pasture Sci.* 64, 361–373. doi: 10.1071/CP13092
- Bolte, S., and Cordelières, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* 224, 213–232. doi: 10.1111/j.1365-2818.2006.01706.x
- Bolwell, G. P., Butt, V. S., Davies, D. R., and Zimmerlin, A. (1995). The origin of the oxidative burst in plants. *Free Radic. Res.* 23, 517–532. doi: 10.3109/10715769509065273
- Bolwell, G. P., and Wojtaszek, P. (1997). Mechanisms for the generation of reactive oxygen species in plant defence: a broad perspective. *Physiol. Mol. Plant Pathol.* 51, 347–366. doi: 10.1006/pmpp.1997.0129
- Cândido, E. S., Pinto, M. F. S., Pelegrini, P. B., Lima, T. B., Silva, O. N., Pogue, R., et al. (2011). Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms. *FASEB J.* 25, 3290–3305. doi: 10.1096/fj.11-184291
- Cessna, S. G., Sears, V. E., Dickman, M. B., and Low, P. S. (2000). Oxalic acid, a pathogenicity factor for *Sclerotinia sclerotiorum*, suppresses the oxidative burst of the host plant. *Plant Cell* 12, 2191–2200. doi: 10.2307/3871114
- Chrispeels, M. J., and Raikhel, N. V. (1991). Lectins, lectin genes, and their role in plant defense. *Plant Cell* 3, 1–9. doi: 10.2307/3869195
- Coda, R., Rizzello, C. G., Nigro, F., De Angelis, M., Arnault, P., and Gobetti, M. (2008). Long-term fungal inhibitory activity of water-soluble extracts of *Phaseolus vulgaris* cv. Pinto and sourdough lactic acid bacteria during bread storage. *Appl. Environ. Microbiol.* 74, 7391–7398. doi: 10.1128/AEM.01420-08
- De Souza Candido, E., Pinto, M., Barbosa Pelegrini, P., Bergamin Lima, T., Nascimento Silva, O., Pogue, R., et al. (2011). Plant storage proteins with antimicrobial activity: novel insights into plant defense mechanisms. *FASEB J.* 25, 3290–3305. doi: 10.1096/fj.11-184291
- Doke, N. (1985). NADPH-dependent O₂- generation in membrane fractions isolated from wounded potato tubers inoculated with *Phytophthora infestans*. *Physiol. Plant Pathol.* 27, 311–322. doi: 10.1016/0048-4059(85)90044-X
- Donaldson, P. A., Anderson, T., Lane, B. G., Davidson, A. L., and Simmonds, D. H. (2001). Soybean plants expressing an active oligomeric oxalate oxidase from the wheat gf-2.8 (germin) gene are resistant to the oxalate-secreting pathogen *Sclerotinia sclerotiorum*. *Physiol. Mol. Plant Pathol.* 59, 297–307. doi: 10.1006/pmpp.2001.0369
- Dong, X., Ji, R., Guo, X., Foster, S. J., Chen, H., Dong, C., et al. (2008). Expressing a gene encoding wheat oxalate oxidase enhances resistance to *Sclerotinia sclerotiorum* in oilseed rape (*Brassica napus*). *Planta* 228, 331–340. doi: 10.1007/s00425-008-0740-2
- Dunwell, J. M., Culham, A., Carter, C. E., Sosa-Aguirre, C. R., and Goodenough, P. W. (2001). Evolution of functional diversity in the cupin superfamily. *Trends Biochem. Sci.* 26, 740–746. doi: 10.1016/S0968-0004(01)01981-8
- Dunwell, J. M., Purvis, A., and Khuri, S. (2004). Cupins: the most functionally diverse protein superfamily? *Phytochemistry* 65, 7–17. doi: 10.1016/j.phytochem.2003.08.016
- Duranti, M., and Gius, C. (1997). Legume seeds: protein content and nutritional value. *Field Crops Res.* 53, 31–45. doi: 10.1016/S0378-4290(97)00021-X
- Fischer, K., Dieterich, R., Nelson, M. N., Kamphuis, L. G., Singh, K. B., Rotter, B., et al. (2015). Characterization and mapping of LanrBo: a locus conferring anthracnose resistance in narrow-leaved lupin (*Lupinus angustifolius* L.). *Theor. Appl. Genet.* 128, 2121–2130. doi: 10.1007/s00122-015-2572-3
- Foley, R. C., Gao, L. L., Spriggs, A., Soo, L. Y., Goggins, D. E., Smith, P. M., et al. (2011). Identification and characterisation of seed storage protein transcripts from *Lupinus angustifolius*. *BMC Plant Biol.* 11:59. doi: 10.1186/1471-2229-11-59
- Foley, R. C., Jimenez-Lopez, J. C., Kamphuis, L. G., Hane, J. K., Melser, S., and Singh, K. B. (2015). Analysis of conglutinin seed storage proteins across lupin species using transcriptomic, protein and comparative genomic approaches. *BMC Plant Biol.* 15:106. doi: 10.1186/s12870-015-0485-6

FIGURE S4 | Recombinant β 1-conglutinin exhibits *in planta* anti-fungal and oomycete activity. Shown are representative images of *Agrobacterium* infiltrated *N. benthamiana* leaves expressing recombinant β 1-conglutinin proteins and subsequently inoculated with either *S. sclerotiorum* or *P. nicotianae*. The experiment was repeated three times with similar results.

FIGURE S5 | Recombinant β 1-conglutinin reduces pathogen growth and pathogen induced cell death *in planta*. Shown are representative images of *Agrobacterium* infiltrated *N. benthamiana* leaves expressing recombinant β 1-conglutinin proteins and subsequently inoculated either *S. sclerotiorum* or *P. nicotianae*. Trypan blue staining was performed to visualize hyphal growth and cell death. Arrows point to hyphae, asterisk marks inoculation site.

FIGURE S6 | β 1-conglutinin is localized to the cell surface and plasmodesmata. (A) Confocal images of tobacco epidermis cell expressing GFP alone or β 1-GFP. Insert: β 1-GFP shows punctate labeling at the cell surface. **(B–E)** Single-slice confocal images of co-expression β 1-GFP with the plasmodesmata marker PPDLP1-mCherry after transient expression in *N. benthamiana*; **(B)** PDL1-mCherry, **(C)** β 1-GFP, **(D)** Image showing pixel pairs that have a positive PDM value equal to the value (intensity of B – mean B intensity) * (intensity of C – mean C intensity) as described in Li et al. (2004), **(E)** merge of **(B,C)** with highlighted co-localized pixels. **ICQ**, Intensity correlation quotient; **R** = Mandel's overlap coefficient. 60 \times immersion objective.

- Gábrisová, D., Klubicová, K., Danchenko, M., Gömöry, D., Berezina, V. V., Skultety, L., et al. (2016). Do cupins have a function beyond being seed storage proteins? *Front. Plant Sci.* 6:1215. doi: 10.3389/fpls.2015.01215
- Gallup, C. A., Sullivan, M. J., and Shew, H. D. (2006). Black shank of tobacco. *Plant Health Instructor* doi: 10.1094/PHI-I-2006-0717-01
- Gapper, C., and Dolan, L. (2006). Control of plant development by reactive oxygen species. *Plant Physiol.* 141, 341–345. doi: 10.1104/pp.106.079079
- Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* 43, 205–227. doi: 10.1146/annurev.phyto.43.040204.135923
- Gomes, L. S., Senna, R., Sandim, V., Silva-Neto, M. A. C., Perales, J. E. A., Zingali, R. B., et al. (2014). Four conventional soybean [*Glycine max* (L.) Merrill] seeds exhibit different protein profiles as revealed by proteomic analysis. *J. Agric. Food Chem.* 62, 1283–1293. doi: 10.1021/jf404351g
- Gomes, V. M., Mosqueda, M.-I., Blanco-Labra, A., Sales, M. P., Fernandes, K. V. S., Cordeiro, R. A., et al. (1997). Vicilin storage proteins from *Vigna unguiculata* (Legume) seeds inhibit fungal growth. *J. Agric. Food Chem.* 45, 4110–4115. doi: 10.1021/jf960942g
- Gomes, V. M., Okorokov, L. A., Rose, T. L., Fernandes, K. V., and Xavier-Filho, J. (1998). Legume vicilins (7S storage globulins) inhibit yeast growth and glucose stimulated acidification of the medium by yeast cells. *Biochim. Biophys. Acta* 1379, 207–216. doi: 10.1016/S0304-4165(97)00100-1
- Ham, B. K., Li, G., Kang, B. H., Zeng, F., and Lucas, W. J. (2012). Overexpression of *Arabidopsis* plasmodesmata germin-like proteins disrupts root growth and development. *Plant Cell* 24, 3630–3648. doi: 10.1105/tpc.112.101063
- Hane, J. K., Anderson, J. P., Williams, A. H., Sperschneider, J., and Singh, K. B. (2014). Genome sequencing and comparative genomics of the broad host-range pathogen *Rhizoctonia solani* AG8. *PLoS Genet.* 10:e1004281. doi: 10.1371/journal.pgen.1004281
- Hane, J. K., Ming, Y., Kamphuis, L. G., Nelson, M. N., Garg, G., Atkins, C. A., et al. (2016). A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant Biotechnol. J.* doi: 10.1111/pbi.12615 [Epub ahead of print].
- Jimenez-Lopez, J. C., Lima-Cabello, E., Melser, S., Foley, R. C., Singh, K. B., and Alche, J. D. (2015). Lupin allergy: uncovering structural features and epitopes of β -conglutin proteins in *Lupinus angustifolius* L. with a focus on cross-allergenic reactivity to peanut and other legumes. *Lect. Notes Comput. Sci.* 9043, 96–107. doi: 10.1007/978-3-319-16483-0_10
- Jimenez-Lopez, J. C., Zienkiewicz, A., Zienkiewicz, K., Alche, J. D., and Rodriguez-Garcia, M. I. (2016). Biogenesis of protein bodies during legumin accumulation in *Olea europaea* L. seed development and differentiation. *Protoplasma* 253, 517–530. doi: 10.1007/s00709-015-0830-5
- Jin, T., Wang, Y., Chen, Y. W., Fu, T. J., Kothary, M. H., McHugh, T. H., et al. (2014). Crystal structure of Korean pine (*Pinus koraiensis*) 7S seed storage protein with copper ligands. *J. Agric. Food Chem.* 62, 222–228. doi: 10.1021/jf4039887
- Kamphuis, L. G., Hane, J. K., Nelson, M. N., Gao, L., Atkins, C. A., and Singh, K. B. (2015). Transcriptome sequencing of different narrow-leaved lupin tissue types provides a comprehensive uni-gene assembly and extensive gene-based molecular markers. *Plant Biotechnol. J.* 13, 14–25. doi: 10.1111/pbi.12229
- Kärkönen, A., and Kuchitsu, K. (2015). Reactive oxygen species in cell wall metabolism and development in plants. *Phytochemistry* 112, 22–32. doi: 10.1016/j.phytochem.2014.09.016
- Keogh, R. C., Deverall, B. J., and McLeod, S. (1980). Comparison of histological and physiological responses to *Phakopsora pachyrhizi* in resistant and susceptible soybean. *Trans. Br. Mycol. Soc.* 74, 329–333. doi: 10.1016/S0007-1536(80)80163-X
- Khuri, S., Bakker, F. T., and Dunwell, J. M. (2001). Phylogeny, function, and evolution of the cupins, a structurally conserved, functionally diverse superfamily of proteins. *Mol. Biol. Evol.* 18, 593–605. doi: 10.1093/oxfordjournals.molbev.a003840
- Kim, H. J., Chen, C., Kabbage, M., and Dickman, M. B. (2011). Identification and characterization of *Sclerotinia sclerotiorum* NADPH oxidases. *Appl. Environ. Microbiol.* 77, 7721–7729. doi: 10.1128/AEM.05472-11
- Kim, K. S., Min, J. Y., and Dickman, M. B. (2008). Oxalic acid is an elicitor of plant programmed cell death during *Sclerotinia sclerotiorum* disease development. *Mol. Plant Microbe Interact.* 21, 605–612. doi: 10.1094/MPMI-21-5-0605
- Kukavica, B., Vucinic, Z., and Vuletic, M. (2005). Superoxide dismutase, peroxidase, and germin-like protein activity in plasma membranes and apoplast of maize roots. *Protoplasma* 226, 191–197. doi: 10.1007/s00709-005-0112-8
- Laluk, K., and Mengiste, T. (2010). Necrotroph attacks on plants: wanton destruction or covert extortion? *The Arabidopsis Book* 8:e0136. doi: 10.1199/tab.0136
- Lane, B. G., Cuming, A. C., Frégeau, J., Carpita, N. C., Hurkman, W. J., Bernier, F., et al. (1992). Germin isoforms are discrete temporal markers of wheat development. Pseudogermin is a uniquely thermostable water-soluble oligomeric protein in ungerminated embryos and like germin in germinated embryos, it is incorporated into cell walls. *Eur. J. Biochem.* 209, 961–969. doi: 10.1111/j.1432-1033.1992.tb17369.x
- Lee, J. Y., and Lu, H. (2011). Plasmodesmata: the battleground against intruders. *Trends Plant Sci.* 16, 201–210. doi: 10.1016/j.tplants.2011.01.004
- Levine, A., Tenhaken, R., Dixon, R., and Lamb, C. (1994). H₂O₂ from the oxidative burst orchestrates the plant hypersensitive disease resistance response. *Cell* 79, 583–593. doi: 10.1016/0092-8674(94)90544-4
- Li, Q., Lau, A., Morris, T. J., Guo, L., Fordyce, C. B., and Stanley, E. F. (2004). A syntaxin 1, Galpha(o), and N-type calcium channel complex at a presynaptic nerve terminal: analysis by quantitative immunocolocalization. *J. Neurosci.* 24, 4070–4081. doi: 10.1523/JNEUROSCI.0346-04.2004
- Liu, H., Ma, X., Yu, H., Fang, D., Li, Y., Wang, X., et al. (2016). Genomes and virulence difference between two physiological races of *Phytophthora nicotianae*. *Gigascience* 5:3. doi: 10.1186/s13742-016-0108-7
- Ma, L., Lukasik, E., Gawehns, F., and Takken, F. L. W. (2012). The use of agroinfiltration for transient expression of plant resistance and fungal effector proteins in *Nicotiana benthamiana* leaves. *Methods Mol. Biol.* 835, 61–74. doi: 10.1007/978-1-61779-501-5_4
- Marcus, J. P., Green, J. L., Goulter, K. C., and Manners, J. M. (1999). A family of antimicrobial peptides is produced by processing of a 7S globulin protein in *Macadamia integrifolia* kernels. *Plant J.* 19, 699–710. doi: 10.1046/j.1365-313x.1999.00569.x
- Melo, T. S., Ferreira, R. B., and Teixeira, A. N. (1994). The seed storage proteins from *Lupinus albus*. *Phytochemistry* 37, 641–648. doi: 10.1016/S0031-9422(00)90331-5
- Mittler, R. (2002). Oxidative stress, antioxidants and stress tolerance. *Trends Plant Sci.* 7, 405–410. doi: 10.1016/S1360-1385(02)02312-9
- Molla, K. A., Karmakar, S., Chanda, P. K., Ghosh, S., Sarkar, S. N., Dattam, S. K., et al. (2013). Rice oxalate oxidase gene driven by green tissue-specific promoter increases tolerance to sheath blight pathogen (*Rhizoctonia solani*) in transgenic rice. *Mol. Plant Pathol.* 14, 910–922. doi: 10.1111/mpp.12055
- Monteiro, S., Carreira, A., Freitas, R., Pinheiro, A. M., and Ferreira, R. B. (2015). A nontoxic polypeptide oligomer with a fungicide potency under agricultural conditions which is equal or greater than that of their chemical counterparts. *PLoS ONE* 10:e0122095. doi: 10.1371/journal.pone.0122095
- Morscher, F., Kranner, I., Arc, E., Bailly, C., and Roach, T. (2015). Glutathione redox state, tocopherols, fatty acids, antioxidant enzymes and protein carbonylation in sunflower seed embryos associated with after-ripening and ageing. *Ann. Bot.* 116, 669–678. doi: 10.1093/aob/mcv108
- Murray, G. M., and Brennan, J. P. (2012). *The Current and Potential Costs from Diseases of Pulse Crops in Australia: GRDC Research Code: CER00002*. Barton, ACT: Grains Research and Development Corporation.
- Niyonsaba, F., Nagaoka, I., Ogawa, H., and Okumura, K. (2009). Multifunctional antimicrobial proteins and peptides: natural activators of immune systems. *Curr. Pharm. Des.* 15, 2393–2413. doi: 10.2174/138161209788682271
- Oliveira, A. E. A., Gomes, V. M., Sales, M. P., Fernandes, K. V. S., Carlini, C. R., and Xavier-Filho, J. (1999). The toxicity of jack bean [*Canavalia ensiformis* (L.) dc.] canatoxin to plant pathogenic fungi. *Rev. Bras. Biol.* 59, 59–62. doi: 10.1590/S0034-71081999000100008
- Oracz, K., El-Maarouf Bouteau, H., Farrant, J. M., Cooper, K., Belghazi, M., Job, C., et al. (2007). ROS production and protein oxidation as a novel mechanism for seed dormancy alleviation. *Plant J.* 50, 452–465. doi: 10.1111/j.1365-313X.2007.03063.x
- Overvoorde, P. J., Chao, W. S., and Grimes, H. D. (1997). A plasma membrane sucrose-binding protein that mediates sucrose uptake shares structural and sequence similarity with seed storage proteins but remains functionally distinct. *J. Biol. Chem.* 272, 15898–15904. doi: 10.1074/jbc.272.25.15898

- Pan, H. Y., Whittaker, M. M., Bouveret, R., Berna, A., Bernier, F., and Whittaker, J. W. (2007). Characterization of wheat germin (oxalate oxidase) expressed by *Pichia pastoris*. *Biochem. Biophys. Res. Commun.* 356, 925–929. doi: 10.1016/j.bbrc.2007.03.097
- Perl-Treves, R., Foley, R. C., Chen, W., and Singh, K. B. (2004). Early induction of the *Arabidopsis* GSTF8 promoter by specific strains of the fungal pathogen *Rhizoctonia solani*. *Mol. Plant Microbe Interact.* 17, 70–80. doi: 10.1094/MPMI.2004.17.1.70
- Petrie, J. R., Shrestha, P., Liu, Q., Mansour, M. P., Wood, C. C., Zhou, X.-R., et al. (2010). Rapid expression of transgenes driven by seed-specific constructs in leaf tissue: DHA production. *Plant Methods* 6, 1–6. doi: 10.1186/1746-4811-6-8
- Rietz, S., Bernsdorff, F. E., and Cai, D. (2012). Members of the germin-like protein family in *Brassica napus* are candidates for the initiation of an oxidative burst that impedes pathogenesis of *Sclerotinia sclerotiorum*. *J. Exp. Bot.* 63, 5507–5519. doi: 10.1093/jxb/ers203
- Rinalducci, S., Murgiano, L., and Zolla, L. (2008). Redox proteomics: basic principles and future perspectives for the detection of protein oxidation in plants. *J. Exp. Bot.* 59, 3781–3801. doi: 10.1093/jxb/ern252
- Rubiales, D., Fondevilla, S., Chen, W., Gentzittel, L., Higgins, T. J. V., Castillejo, M. A., et al. (2015). Achievements and challenges in legume breeding for pest and disease resistance. *CRC Crit. Rev. Plant Sci.* 34, 1–3. doi: 10.1080/07352689.2014.898445
- Sales, M. P., Pimenta, P. P., Paes, N. S., Grossi-de-Sá, M. F., and Xavier-Filho, J. (2001). Vicilins (7S storage globulins) of cowpea (*Vigna unguiculata*) seeds bind to chitinous structures of the midgut of *Callosobruchus maculatus* (Coleoptera: Bruchidae) larvae. *Braz. J. Med. Biol. Res.* 34, 27–34. doi: 10.1590/S0100-879X2001000100003
- Scarafoni, A., Ronchi, A., Prinsi, B., Espen, L., Assante, G., Venturini, G., et al. (2013). The proteome of exudates from germinating *Lupinus albus* seeds is secreted through a selective dual-step process and contains proteins involved in plant defence. *FEBS J.* 280, 1443–1459. doi: 10.1111/febs.12140
- Schenk, P. M., Kazan, K., Wilson, I., Anderson, J. P., Richmond, T., Somerville, S. C., et al. (2000). Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11655–11660. doi: 10.1073/pnas.97.21.11655
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Schopfer, P., and Liskay, A. (2006). Plasma membrane-generated reactive oxygen intermediates and their role in cell growth of plants. *Biofactors* 28, 73–81. doi: 10.1002/biof.5520280202
- Schweizer, P., Christoffel, A., and Dudler, R. (1999). Transient expression of members of the germin-like gene family in epidermal cells of wheat confers disease resistance. *Plant J.* 20, 541–552. doi: 10.1046/j.1365-313X.1999.00624.x
- Sewelam, N., Kazan, K., and Schenk, P. M. (2016). Global plant stress signaling: reactive oxygen species at the cross-road. *Front. Plant Sci.* 7:187. doi: 10.3389/fpls.2016.00187
- Shetty, N. P., Jorgensen, H. J. L., Jensen, J. D., Collinge, D. B., and Shetty, H. S. (2008). Roles of reactive oxygen species in interactions between plants and pathogens. *Eur. J. Plant Pathol.* 121, 267–280. doi: 10.1007/s10658-008-9302-5
- Sinden, J. A., Jones, R., Hester, S., Odom, D., Kalisch, C., James, R., et al. (2004). *The Economic Impact of Weeds in Australia: Report to the CRC for Australian Weed Management*. Glen Osmond, SA: CRC for Australian Weed Management.
- Sparkes, I. A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* 1, 2019–2025. doi: 10.1038/nprot.2006.286
- Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Prot. Exp. Pur.* 41, 207–234. doi: 10.1016/j.pep.2005.01.016
- Tam, J. P., Wang, S., Wong, K. H., and Tan, W. L. (2015). Antimicrobial peptides from plants. *Pharmaceuticals* 8, 711–757. doi: 10.3390/ph8040711
- Tan-Wilson, A. L., and Wilson, K. A. (2012). Mobilization of seed protein reserves. *Physiol. Plant.* 145, 140–153. doi: 10.1111/j.1399-3054.2011.01535.x
- Thatcher, L. F., Anderson, J. P., and Singh, K. B. (2005). Plant defence responses: what have we learnt from *Arabidopsis*? *Funct. Plant Biol.* 32, 1–19. doi: 10.1071/FP04135
- Thatcher, L. F., Carrie, C., Andersson, C. R., Sivasithamparan, K., Whelan, J., and Singh, K. B. (2007). Differential gene expression and subcellular targeting of *Arabidopsis* glutathione S-transferase F8 is achieved through alternative transcription start sites. *J. Biol. Chem.* 282, 28915–28928. doi: 10.1074/jbc.M702207200
- Thatcher, L. F., Gardiner, D. M., Kazan, K., and Manners, J. M. (2012). A highly conserved effector in *Fusarium oxysporum* is required for full virulence on *Arabidopsis*. *Mol. Plant Microbe Interact.* 25, 180–190. doi: 10.1094/MPMI-08-11-0212
- Thomas, C. L., Bayer, E. M., Ritzenthaler, C., Fernandez-Calvino, L., and Maule, A. J. (2008). Specific targeting of a plasmodesmal protein affecting cell-to-cell communication. *PLoS Biol.* 6:e7. doi: 10.1371/journal.pbio.0060007
- Torres, M. A., Jones, J. D. G., and Dangel, J. L. (2006). Reactive oxygen species signaling in response to pathogens. *Plant Physiol.* 141, 373–378. doi: 10.1104/pp.106.079467
- van Wees, S. (2008). Phenotypic analysis of *Arabidopsis* mutants: trypan blue stain for fungi, oomycetes, and dead plant cells. *CSH Protoc.* 2008:db.rot4982. doi: 10.1101/pdb.prot4982
- Walz, A., Zingen-Sell, I., Loeffler, M., and Sauer, M. (2008). Expression of an oxalate oxidase gene in tomato and severity of disease caused by *Botrytis cinerea* and *Sclerotinia sclerotiorum*. *Plant Pathol.* 57, 453–458. doi: 10.1111/j.1365-3059.2007.01815.x
- Wang, T., Chen, X., Zhu, F., Li, H., Li, L., Yang, Q., et al. (2013). Characterization of peanut germin-like proteins, AhGLPs in plant development and defense. *PLoS ONE* 8:e61722. doi: 10.1371/journal.pone.0061722
- Wang, X., Bunkers, G. J., Walters, M. R., and Thoma, R. S. (2001). Purification and characterization of three antifungal proteins from cheeseweed (*Malva parviflora*). *Biochem. Biophys. Res. Commun.* 282, 1224–1228. doi: 10.1006/bbrc.2001.4716
- Williams, A. H., Sharma, M., Thatcher, L. F., Azam, S., Hane, J. K., Sperschneider, J., et al. (2016). Comparative genomics and prediction of conditionally dispensable sequences in legume-infecting *Fusarium oxysporum* formae speciales facilitates identification of candidate effectors. *BMC Genomics* 17:191. doi: 10.1186/s12864-016-2486-8
- Williams, B., Kabbage, M., Kim, H.-J., Britt, R., and Dickman, M. B. (2011). Tipping the balance: *Sclerotinia sclerotiorum* secreted oxalic acid suppresses host defenses by manipulating the host redox environment. *PLoS Pathog.* 7:e1002107. doi: 10.1371/journal.ppat.1002107
- Wojtaszek, P. (1997). Oxidative burst: an early plant response to pathogen infection. *Biochem. J.* 322, 681–692. doi: 10.1042/bj3220681
- Woo, E. J., Dunwell, J. M., Goodenough, P. W., Marvier, A. C., and Pickersgill, R. W. (2000). Germin is a manganese containing homohexamer with oxalate oxidase and superoxide dismutase activities. *Nat. Struct. Mol. Biol.* 7, 1036–1040. doi: 10.1038/80954
- Wu, G., Shortt, B. J., Lawrence, E. B., Leon, J., Fitzsimmons, K. C., Levine, E. B., et al. (1997). Activation of host defence mechanisms by elevated production of H₂O₂ in transgenic plants. *Plant Physiol.* 115, 427–435. doi: 10.1104/pp.115.2.427
- Yunes, A. N. A., de Andrade, M. T., Sales, M. P., Morais, R. A., Fernandes, K. V. S., Gomes, V. M., et al. (1998). Legume seed vicilins (7S storage proteins) interfere with the development of the cowpea weevil (*Callosobruchus maculatus* (F)). *J. Sci. Food Agric.* 76, 111–116. doi: 10.1002/(SICI)1097-0010(199801)76:1<111::AID-JSFA932>3.0.CO;2-4
- Zienkiewicz, A., Jimenez-Lopez, J. C., Zienkiewicz, K., Alché, J. D., and Rodriguez-Garcia, M. I. (2011). Development of the cotyledon cells during olive (*Olea europaea* L.) in vitro seed germination and seedling growth. *Protoplasma* 248, 751–765. doi: 10.1007/s00709-010-0242-5

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer SF and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Jimenez-Lopez, Melser, DeBoer, Thatcher, Kamphuis, Foley and Singh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Non-host Resistance: DNA Damage Is Associated with SA Signaling for Induction of PR Genes and Contributes to the Growth Suppression of a Pea Pathogen on Pea Endocarp Tissue

Lee A. Hadwiger* and Kiwamu Tanaka

Department of Plant Pathology, Washington State University, Pullman, WA, USA

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Consejo Superior de Investigaciones
Científicas, Spain

Reviewed by:

Seonghee Lee,
University of Florida, USA
Paola Leonetti,
Consiglio Nazionale Delle Ricerche,
Italy

*Correspondence:

Lee A. Hadwiger
chitosan@wsu.edu

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 22 December 2016

Accepted: 14 March 2017

Published: 04 April 2017

Citation:

Hadwiger LA and Tanaka K (2017)
Non-host Resistance: DNA Damage
Is Associated with SA Signaling
for Induction of PR Genes
and Contributes to the Growth
Suppression of a Pea Pathogen on
Pea Endocarp Tissue.
Front. Plant Sci. 8:446.
doi: 10.3389/fpls.2017.00446

Salicylic acid (SA) has been reported to induce plant defense responses. The transcriptions of defense genes that are responsible for a given plant's resistance to an array of plant pathogens are activated in a process called non-host resistance. Biotic signals capable of carrying out the activation of pathogenesis-related (PR) genes in pea tissue include fungal DNase and chitosan, two components released from *Fusarium solani* spores that are known to target host DNA. Recent reports indicate that SA also has a physical affinity for DNA. Here, we report that SA-induced reactive oxygen species release results in fragment alterations in pea nuclear DNA and cytologically detectable diameter and structural changes in the pea host nuclei. Additionally, we examine the subsequent SA-related increase of resistance to the true pea pathogen *F. solani* f.sp. *pisi* and the accumulation of the phytoalexin pisatin. This is the first report showing that SA-induced PR gene activation may be attributed to the host pea genomic DNA damage and that at certain concentrations, SA can be temporally associated with subsequent increases in the defense response of this legume.

Keywords: non-host resistance, DNA damage, salicylic acid, PR genes, *Fusarium solani*

INTRODUCTION

The salicylic acid (SA) signal receptor protein NPR1 has been reported in *Arabidopsis* (Wu et al., 2012), and NPR1 is a known link between SA signaling and defense gene activation. An alternate hypothesis for signal reception in the legume, pea, indicates that host cell chromatin can both serve as a receptor (Hadwiger, 2015a) and provide the site for increased transcription of pathogenesis-related (PR) genes (Isaac et al., 2009). DNA damage within chromatin can also initiate signaling cascades in animal tissues (De Dieuleveult et al., 2016) and is dependent on ubiquitin (Stewart et al., 2009). In rice and peas, chromatin changes can result in the suppression of innate immunity (Li et al., 2015) or the enhancement of PR gene transcription (Chang et al., 1995; Hadwiger, 2009, 2015b; Isaac et al., 2009), respectively. Recent reports (Neauault et al., 1996; Yan et al., 2013) indicate that SA has an affinity for DNA, suggesting the potential of a DNA target site for SA that may add to or supersede reception by a cytoplasmic protein. The model pea endocarp/bean pathogen interaction system is a suitable system to research the role of SA in non-host defense in legumes.

Non-host resistance differs as it is more durable than the single dominant resistance genes commonly manipulated by plant breeders. However, both mechanisms are associated with the enhanced synthesis of PR proteins that are usually involved in plant defense (Klosterman et al., 2001). Genes for many of the PR proteins have been cloned (van Loon, 1985), and their antifungal properties have been identified, e.g., PR2, β -glucanases; PR3 and PR4, chitinases; PR5, thaumatin-like proteins; PR6, proteinase inhibitors; PR7, endoproteases; PR8, cucumber chitinase III; PR9 peroxidase; PR10 ribonuclease-like; PR11, chitinase V; and PR12, defensin; PR13, thionin; PR14, lipid transfer proteins; and PR15 and PR16, oxalate oxidases (van Loon, 1985). Additionally, many of the single dominant genes (R genes) identified in diverse collections of a given plant species have also been cloned and bred into plants for disease resistance (Presti et al., 2015). The products of the R genes often recognize specific pathogen effectors (Zhou et al., 1997; Boller and Felix, 2009; Hadwiger and Chang, 2015). These genes are efficiently utilized for crop improvement. However, the resistance they provide can be bypassed by mutations in the effector genes of the pathogen (McDowell and Woffenden, 2003).

The non-host resistance that enables plants outside the host range of a given pathogen to resist their “inappropriate” pathogens is probably more durable because there are diverse types of effectors/elicitors and because multiple resistance traits are involved. To account for the multiple effector/plant protein receptor roles in the PAMP/PRR defense model (Boller and Felix, 2009), one must hypothesize a pre-event presence of an abundant gene bank of plant receptor proteins that is broad enough to match all the diverse effectors the plant may confront. This signaling event must also be capable of transmitting the signal to the site for defense gene transcription (Hadwiger, 2015b). The non-host resistance model with chromatin as a receptor offers flexibility to account for many of the multiple interactions between plants and their pathogens. This resistance developed against an “inappropriate” pathogenic fungus, such as a bean pathogen in pea, can rapidly develop within the pea endocarp tissue (Hadwiger, 2015b). Some major receptors targeted by effectors/elicitors released by these fungi may lie directly within the DNA and proteins of pea chromatin (Isaac et al., 2009). There are diverse mechanisms, such as remodeling or altering transcription and enhancing the properties of chromatin (Li et al., 2007), that result in PR gene activation. The multiplicity of DNA conformations or the modifications of the nuclear proteins in plant chromatin have been described (Choi et al., 2001), which include DNA strand breakage, base substitution, helical changes, deletion/point mutations, nuclear protein removal (ubiquitination) and histone modification or elimination, among others (Li et al., 2007; Lagerwerf et al., 2011).

Thus, the objective of the current research was to evaluate the aspects of legume defense simulation by SA (capable of signaling disease resistance in *Arabidopsis*) that may correspond with the induction of non-host resistance by *Fusarium solani* f.sp. *phaseoli* (*Fsph*), an inducer of non-host resistance in pea tissue. This analysis examined the development of reactive oxygen species

(ROS) and DNA damage in pea tissue. Subsequently, the resultant SA-related activation of pea PR genes important to plant defense was monitored with DNA probes from pea genes possessing partial homology to those in *Arabidopsis*.

The molecular response between fungal pathogens and plant cells is rapid if the signaling route excludes surface obstacles, such as the cuticle layer. The pea endocarp system was selected because the entire surface lacks a cuticle, and the surfaces of epidermal cells uniformly react to fungal inoculum, providing total resistance to non-pathogenic or inappropriate pathogens within 6 h. Additionally, the nuclei within the surface cell layer can be easily stained and monitored for visible changes. Time course increases in ROS and changes in DNA fragmentation can be readily assayed to evaluate their participation in initiating the transcription of PR genes, especially those with protein products such as the defensins that directly suppress growth of pathogenic spores (Almeida et al., 2000). Increases in ROS have reportedly been implicated in increasing DNA damage. DNA damage in the pea host is also associated with the release of fungal DNase, a mitochondrial DNase from *Fsph* (Klosterman et al., 2001), thus suggesting that the effects of overlapping DNA damage help to initiate gene transcription.

Pea PR genes map to multiple chromosomes and often reside in regions that also map as QTLs (Pilet-Nayel et al., 2002). PR genes are ubiquitously present in plant genomes and possess properties that enable them to be selectively expressed in the resistance response. PR genes with strong antifungal properties are potentially major contributors to resistance (Chiang and Hadwiger, 1991; Almeida et al., 2000). It appears that there is an additive effect of multiple PR genes that results in complete non-host resistance. Pea PR genes share partial homology with the PR genes induced by SA in *Arabidopsis* (Sels et al., 2008). The objective of this research was to determine whether the genes activated by SA respond similarly to those induced by other elicitors in pea endocarp tissue. An additional objective was to determine whether there is an associated release of ROS in the early hours following SA treatment.

The SA affinity to DNA, similar to other previously described DNA-specific agents, can cause DNA damage (Neault et al., 1996; Yan et al., 2013). More recently, there have been reports of ATP-dependent chromatin remodelers that allow both transcription factors and the general transcription machinery access to DNA. In addition, these remodelers target specific nucleosomes at the edge of nucleosome-free regions, where they regulate specific transcriptional programs. Nucleosome regions have been identified by DNase 1 digestion assays as areas often encompassing unexpressed genes. This somewhat preferential transcription of PR genes gives credence to the observed selective expression of plant defense resulting from general challenges to sensitive chromatin structures. In cells, the double stranded DNA helix is mostly supercoiled and is either under- or overwound (Ma et al., 2013). RNA polymerase II must transcribe through this supercoiled DNA. For transcription to occur, the DNA helix must be opened as the polymerase threads the separated strands through the enzyme. This process generates supercoiling ahead of and behind the polymerase. The upstream torque disrupts the

DNA double strand structure and stalls the polymerase, while the release of this torsional stress allows the polymerase to resume transcription.

DNA damage by microbial enzymes that cause double stranded breaks has also been reported (Song and Bent, 2014), and it is likely that this higher level damage is more of a challenge to the plant than the single strand nicking caused by *Fsph* DNase. Interestingly, the abundance of double strand breaks is reduced by plant defense responses, suggesting that the mechanisms for activating DNA repair processes may share some similarity with the induction of PR genes.

Since SA has recently been reported (Bau et al., 2013) to interact with DNA and has the potential to indirectly influence the state of nuclear DNA by its catalytic inhibition of topoisomerase II, it also has the potential to influence nuclear DNA in plant cells. Single-strand nicks within the large genomic DNA of plants do not produce fragments small enough to be easily detected by typical DNA separations. Therefore, a post-extraction processing of the total DNA was employed to detect the DNA damage occurring in the very early hours of fungal-plant interactions that activate temporally associated defense responses within the host and non-host plant responses (Hadwiger and Adams, 1978). We describe an alkaline buffer treatment protocol that separates the DNA strands. This preparation is incorporated into CHEF gel agar-plug-like disks to entrap the bulk of the plant genomic DNA while allowing shorter fragments, now single stranded, to be released in adjacent alkaline buffer and quantified (Choi et al., 2001). Thus, the extent of host DNA damage could be based on the amount of fragments released. We have observed that SA can target and fragment pea DNA. There was a release of ROS that may additionally serve as a signaling component. The SA-generated signals appeared inefficient at activating the secondary metabolism required to produce maximal amounts of pisatin. The transcription response to the SA and fungal challenges was measured with PCR measurements of alterations in the expression of the selected PR genes.

MATERIALS AND METHODS

Plant Material

Pea endocarp tissue was obtained from immature pea pods harvested directly from greenhouse-grown (Samish) peas. The pod halves were separated, and the elicitor treatments were applied to the exposed endocarp tissue.

Luminol-Based Oxidative Burst Assay

Immature, 2-cm-long pea pods were cut in half. For each sample, one piece (~1 cm in diameter) was immersed in deionized water in a single well of a white 24-well microplate (PerkinElmer). After an overnight incubation, the solution in the well was exchanged with assay solution containing 100 μ M of L-012 (luminol analog; Wako) and 20 μ g/ml of horseradish peroxidase (Sigma-Aldrich), with or without SA. The luminescence from each well was measured using an EnSpire multimode plate reader (PerkinElmer).

Fungal Material

The bean pathogen *F. solani* f.sp. *phaseoli*, Snyder and Hansen (*Fsph*) (ATCC no. 38135) was donated from the Doug Burke lab, and the pea pathogen *F. solani* f.sp. *pisi* (*Fspi*) was obtained from Lindon Porter, IAREC, Prosser, WA.

Plant Nucleic Acid Extraction and Quantitation

Plant tissue was extensively ground in a mortar with liquid N₂, glass beads, and the nucleic acids were extracted in buffer no. 1 [5 M sodium perchlorate, 0.5 M Tris base, 2.5% (w/v) SDS, 0.05% (w/v) NaCl, 0.05 M EDTA]. DNA/RNA were precipitated with 95% (v/v) ethanol, and the pellet was redissolved in water, subsequently extracted with chloroform/phenol, and redissolved in water. The RNA was precipitated from the extract by treating the solution with 2 M lithium chloride. The RNA pellet and the ethanol-precipitated DNA from the supernatant were quantitated in a spectrophotometer at 260 nm. Aliquots of the total DNA were electrophoretically separated on standard 1% (w/v) agarose gels. In addition, 30 μ g of the total of each treatment was incorporated into 1 ml of 1% (w/v) CHEF gel (in a 1.5 diameter well) under alkaline conditions to cause DNA strand separation. The solidified gel disk was overlaid with 1 ml of alkaline buffer (30 ml 1 N NaOH and 8 ml 0.5 M EDTA/L) and rotated for 48 h. The DNA fragments eluted into the overlay were precipitated and separated on standard agarose gels. All of the treatments were repeated with similar results.

Cytological Detection of Treatment-induced Nuclear Changes in Pea Endocarp Tissue

Changes in the nuclear structure and nuclei diameter were imaged with a fluorescent microscope following staining with the DNA-specific dye, DAPI. Subsequently, the diameters of the nuclei from the digital images were uniformly amplified by photocopying, and 45 nuclei from each treatment were manually measured.

Quantitative Real-time RT-PCR (qRT-PCR)

The procedures for the total RNA isolation and purification were performed as described above. The total RNA was subjected to qRT-PCR using a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.) The primers used were described in our previous research (Hadwiger and Tanaka, 2015).

RESULTS AND DISCUSSION

Effect of SA on the Production of ROS in Pea Endocarp Tissue

An “oxidative burst” is the rapid release of ROS from stressed plant cells that develops when they come into contact with different pathogens (Grant and Loake, 2000). To detect this early ROS response in pea, a luminol-based oxidative burst assay was

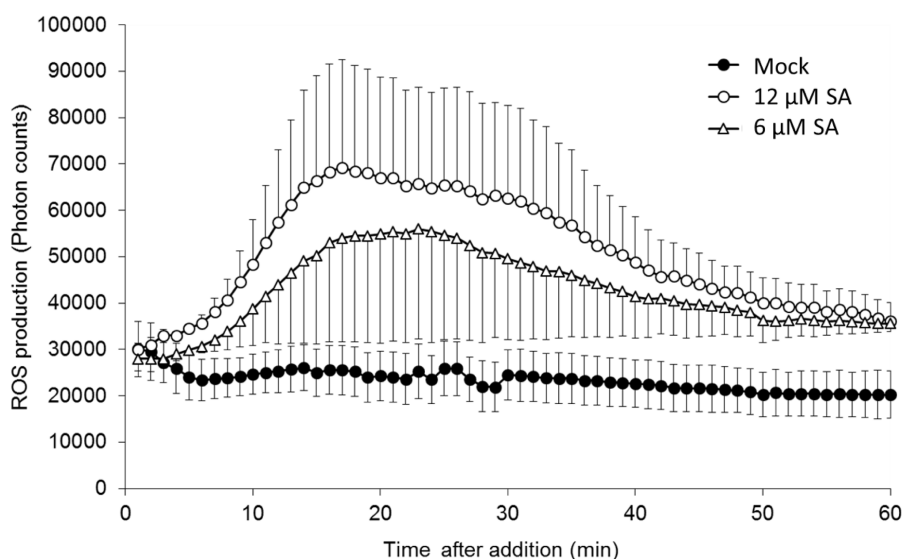


FIGURE 1 | Salicylic acid (SA)-induced reactive oxygen species (ROS) accumulation in pea endocarps. Half-cut pea endocarps were treated without or with 6 or 12 μ M of SA. Treatment with water (Mock) was used as a negative control. The data show the photon counts in 1.0 s at each time point with SE ($n = 4$).

TABLE 1 | Effect of salicylic acid (SA) treatments on the subsequent 24 h growth of *Fspi* on pea endocarp tissue.

Treatment ^a	Concentration	Linear growth of <i>Fspi</i> ^b
SA	100 μ M	2.29 \pm 1.29
SA	50 μ M	2.53 \pm 1.50
SA	25 μ M	1.04 \pm 0.94
SA	12 μ M	0.33 \pm 0.33
SA	6 μ M	1.29 \pm 1.10
SA	3 μ M	0.79 \pm 0.78
SA	1.6 μ M	1.19 \pm 0.49
SA	0.7 μ M	0.85 \pm 0.75
SA	0.3 μ M	2.03 \pm 0.73
SA	0.15 μ M	0.10 \pm 0.10
SA	0.07 μ M	0.62 \pm 0.62
SA	0.03 μ M	1.41 \pm 0.41
Water 1	–	1.83 \pm 1.50
Water 2	–	1.73 \pm 0.43

^aThe endocarp inner tissues of immature pea pod halves (2 cm) were treated with 25 μ l of the indicated treatments. After 20 min, 10 μ l of an *Fspi* suspension (6.7×10^5 spores/ml) was applied to each pod half. The two water treatments used were duplicate.

^bThe growth of 10 individual cotton blue stained spores per treatment was recorded after 24 h. Numbers indicate the multiples of the length of a 45-micron macroconidia.

performed. As shown in **Figure 1**, SA treatment induced an oxidative burst with a peak at ~ 20 min, whereas water treatment (mock) did not induce an oxidative burst. This method is quite robust and sensitively captures the dynamic changes in ROS production at an early time point in the pea endocarp.

Salicylic acid applied 20 min prior to the inoculum at certain concentrations significantly reduced the linear growth of the true pea pathogen *F. solani* f.sp. *pisi* (*Fspi*) on the endocarp

surface. The gradation of action relative to the SA concentration was reproducible over two extensive trials. One of the trials is presented in **Table 1**, while the other is not shown. The growth of the pea pathogen *Fspi* on the pea endocarp surface is less than that on water-treated tissue.

Cytological readings (**Table 1**) of the fungal growth began to demonstrate measurable inhibition after 24 h (**Figure 2**). An SA dilution series treatment down to the 0.03 μ M showed suppressive effects. The characteristic changes in the background hypersensitivity discoloration of the adjacent pea cells suggest that there was a plant-based change in the suppressive effect. Nearly complete and optimal resistance occurred close to the 0.15 μ M SA concentration.

Effect of SA Concentrations on the *In vitro* Growth of the Pea Pathogen *Fspi*

Salicylic acid had no significant direct effect on *in vitro* *Fspi* growth in liquid media. The microscopic examination of growth after 24 h in Vogel's media indicated that the *Fspi* spores germinated and grew uniformly at the concentrations used in **Figure 2** (data not shown).

SA Induced Changes at the DNA/Nuclear Level

Following the report that SA has an affinity for DNA (Neuault et al., 1996), it was of interest to examine changes in pea DNA damage in the nucleus and elsewhere within the pea cells. SA applied to the cuticle-free surface of the pea endocarp tissue rapidly caused cytologically detectable changes in the plant nuclei (**Figure 3** and **Table 2**). These changes were related to the SA concentration and the duration of the SA exposure.

DNA fragmentation appeared rapidly (50 min post treatment) and variably with the range of SA treatment concentrations

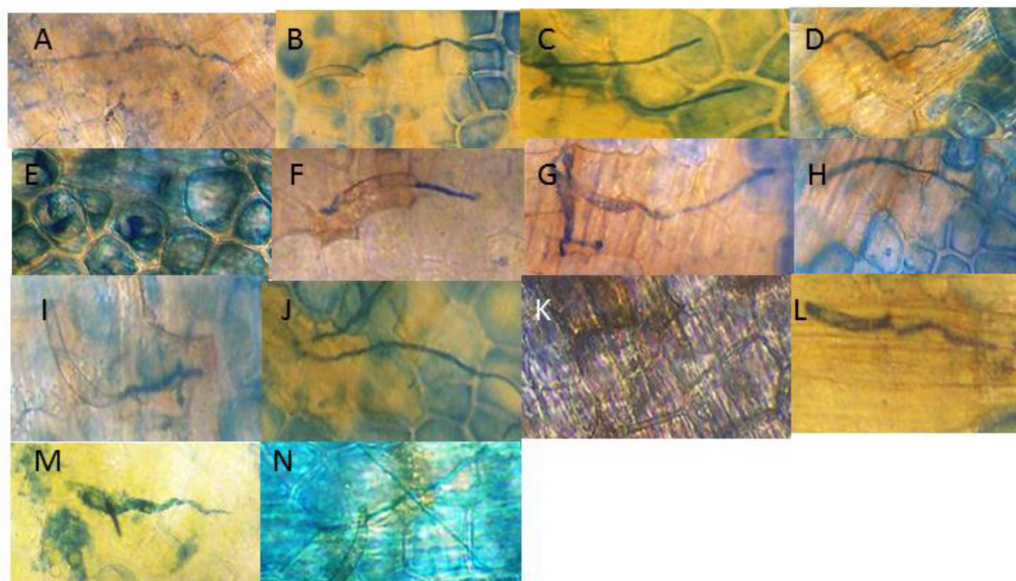


FIGURE 2 | Visual effect of SA treatments of pea endocarp tissue on the 24 h growth of the true pea pathogen *Fsp*. The images are representative of the calculated growth of the SA treatments (indicated in **Table 1**): (A) H₂O; (B) 100 μ M SA; (C) 50 μ M SA; (D) 25 μ M SA; (E) 12 μ M SA; (F) 6 μ M SA; (G) 3 μ M SA; (H) 1.6 μ M SA; (I) 0.7 μ M SA; (J) 0.3 μ M SA; (K) 0.15 μ M SA; (L) 0.07 μ M SA; (M) 0.03 μ M SA; and (N) 0.015 μ M SA. The action of SA for causing DNA damage and enhancing resistance appears variable in regions of the SA gradient. This variability may also correspond with its differing resistance action.

and was generally consistent throughout multiple experiments (**Figure 4**). Fragmentation was more intense for the treatment with 100 to 6.75 μ M SA and for tissues treated with *Fsph* spores. The specific mechanistic impact of SA on the pea DNA responsible for initiating chromatin transcription is not known for either pea or animal tissues. Maximal transcription of PR genes may depend on a “perfect storm” of conditions and the fragility of chromosomal regions adjacent to the promoter and open reading frame of the gene. Regions of dispersed pea chromatin that are also regions of intense transcription have been detected by electron microscopy (Hadwiger, 2015a) as resistance is developing. Genes within eukaryotic tissues can possess the requisite transcription complex with the proper transcription factors in place and still be silent or stalled (Li et al., 2007). We suggest that there may be stalled PR genes that are activated following major DNA or chromatin structural changes caused by the non-specific SA insults within the adjacent regions.

The reported interaction of SA and DNA did not cause major changes to plasmid DNA (**Figure 5**). There were detectable, faster migrating DNA molecules generated at the highest SA concentrations. How these minor changes would reflect on the structure of DNA incorporated into pea chromatin is not known. This result may indicate that the DNA fragmentation caused by SA in living tissue could involve additional components.

Effect of SA on Expression of Pea PR Genes

Induction of PR genes is correlated with the activation of plant defense. We measured the transcriptional induction of the pea PR

genes, *DRR206*, *Defensin*, *PR10*, and *PR1b* in the presence of SA. The results indicate that the expression of the PR genes induced by SA took place mostly at concentrations between 1.5 and 50 μ M (**Figure 6**). The induction levels were comparable to those caused by *Fsph*.

Elicitation of Pisatin

The elicitation of pisatin, a phytoalexin, serves to indicate the activity of a series of secondary metabolism enzymes from phenylalanine through phenylpropanoid structures to isoflavonoid and other phenolics, many of which have fungal-suppressive properties (Bailey and Mansfield, 1982). Pisatin accumulation is often associated with the induction of immunity in peas (Hadwiger, 2008). The data in **Table 3**, with a high SA concentration range (15–1000 μ M), and in **Table 4**, with a lower range (0.7–100 μ M), recorded at 24 h indicate detectable levels of SA-induced pisatin. The response with both ranges indicates a much lower pisatin accumulation than that induced by the intact microconidia of *Fsph* during the authentic non-host resistance response. The 1.5 μ M SA treatment optimally induced pisatin. However, this value is much lower than the level induced by spores. This result suggests that SA is not a major elicitor of this secondary metabolism route of defense responses in pea at this or higher concentrations of the SA elicitor (**Table 4**).

SA Signal: Complete or Additive Effect on Resistance

The low-level effect of SA on phytoalexin synthesis indicates a departure from the mechanisms of other signals for non-host

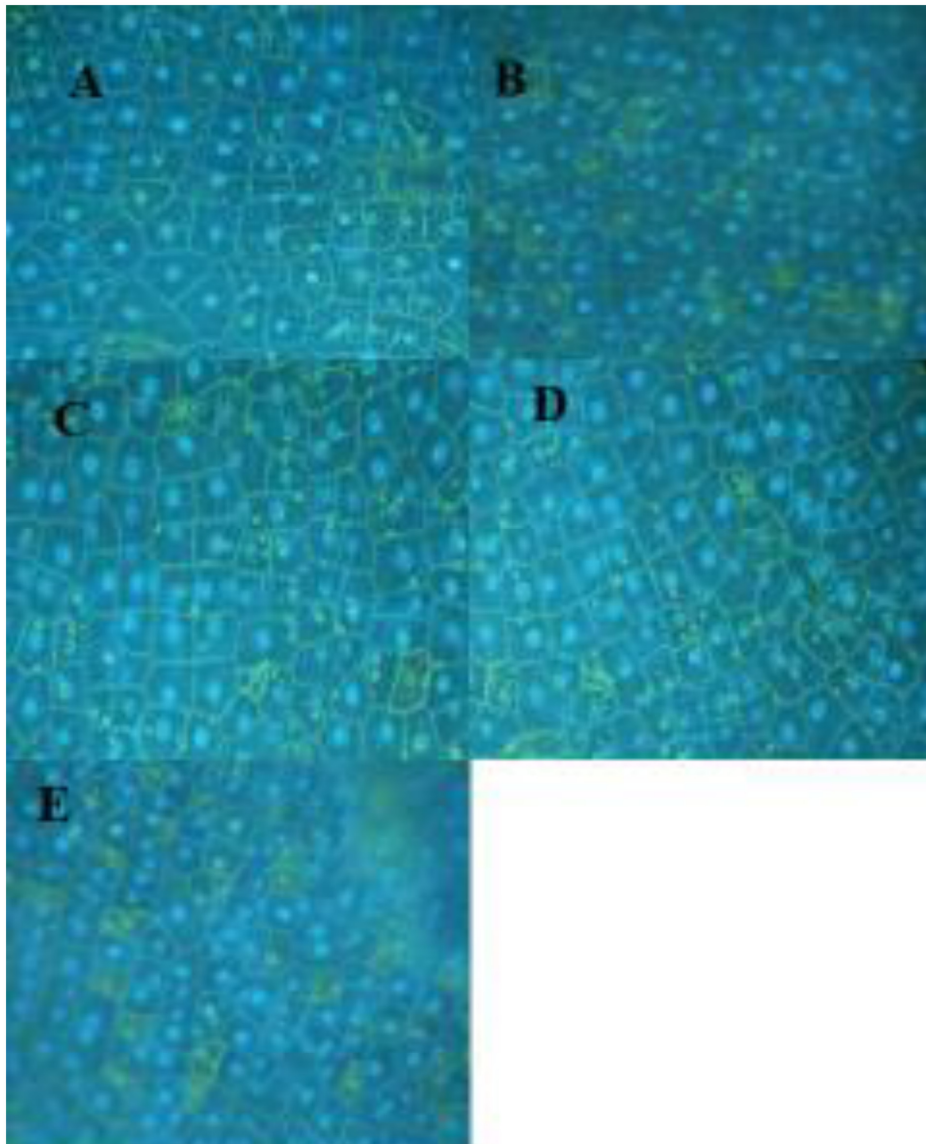


FIGURE 3 | Pea nuclei treated for 3 h with varying levels of SA. **(A)** Water-treated control; **(B)** 1 mM SA; **(C)** 0.5 mM SA; **(D)** 0.125 mM SA; and **(E)** 0.06 mM. Nuclei were stained with DAPI.

resistance in pea. However, SA is capable of inducing a response that suppresses the true pathogen of pea and approaches total resistance. The following assay of pisatin production indicates that its effect can be additive to that induced by *Fsph*, a bean pathogen.

The pisatin levels (Table 5) indicate a marginal increase in synthesis enhanced in the presence of both *Fsph* and specific SA concentrations. Because of the low strength of the SA-induced pisatin increase, it is likely that the modeling effect of SA on chromatin differs in approach or substance from the DNA single strand cleavage generated by *Fsph* DNase (Klosterman, et al., 2001).

The enzymatic action of DNase has also been implicated in initiating the transcription of plant defense genes by

directly altering nuclear chromatin via single DNA strand nicking. The resultant DNA damage has to be subtle enough to alter chromatin structure in a manner that benefits the pathogen and yet does not initiate processes that could cause immediate cell death (Choi et al., 2001). DNA damage by microbial enzymes that cause double stranded breaks has also been reported (Song and Bent, 2014), and it is likely that this higher level damage is more of a challenge to the plant than the single strand nicking caused by *Fsph* DNase. Interestingly, the abundance of double strand breaks is reduced by plant defense responses, suggesting that the mechanisms for activating DNA repair processes may share some similarity with the induction of PR genes (Gasser et al., 2005; Yan et al., 2013).

TABLE 2 | Diameter of nuclei visible in the endocarp surface following treatment with SA dilutions for 30 min.

Treatment	Concentration applied	Diameter ^a Average (μm) of 30 nuclei
Water	—	10.000
SA	100 μM	11.792
SA	50 μM	12.669
SA	25 μM	13.774
SA	12.5 μM	9.630
SA	6.25 μM	10.014
SA	3.12 μM	8.738
SA	1.56 μM	11.522
SA	0.78 μM	8.534

^aNuclei were stained for 5 min with DAPI, and the unfixed tissue was imaged under UV light using a fluorescence microscope. Digital images were uniformly printed on full pages, and 30 nuclei were physically measured; the diameters are compared to that of the water-treated control values and standardized to the 10 micron diameter typically observed in electron micrographs.

Origin of the SA Signal

Some current possibilities for the origin, presence, and availability of the SA signal are described in **Figure 7**. SA is synthesized by bacteria and some fungi (Harper and Hamilton, 1988). SA and methyl-SA can be found in the plant tissue prior to infection and be stored as a byproduct (Maeda and Dudareva, 2012). Hydrogen peroxide is generated in inoculated plant tissue (Coquoz et al., 1988) as tissue damage occurs. In tomatoes, the wound hormone systemin is also produced (Orozco-Cardenas and Ryan, 1999). Hydrogen peroxide can also generate increases in SA. Plants biosynthesize SA using the phenylalanine/cinnamic acid pathway or alternately via benzoic acid (Chen et al., 2009). Both hydrogen peroxide and SA are capable of damaging host DNA. Fungal

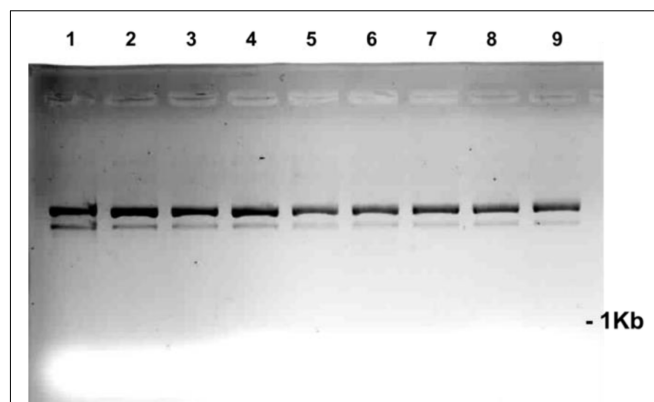


FIGURE 5 | The direct effect of an array of SA concentrations on the migration of pART7F plasmid DNA. The plasmid DNA (0.2 μg in 1 μl) was incubated for 1 h in 5 μl of the SA concentrations in the numbered wells: 1 = 1000 μM; 2 = 100 μM; 3 = 50 μM; 4 = 25 μM; 5 = 12.5 μM; 6 = 6.25 μM; 7 = 3.1 μM; and 8 = 1.5 μM. Well 9 contained only plasmid.

DNase can directly cleave a single DNA strand. The gene for this potent elicitor has been identified in all fungi whose DNA has been sequenced (Hadwiger and Polashock, 2013). All the DNase proteins are translated with a “signalP peptide” that enables proteins to pass through membranes. Many other eliciting components may be released from fungi, such as the chitosan heptamer that is released from the fungal cell wall (Kendra et al., 1989).

Since SA has recently been reported (Bau et al., 2013) to interact with DNA and has the potential to indirectly influence the state of nuclear DNA by its catalytic inhibition of topoisomerase II, it has the potential to also influence nuclear

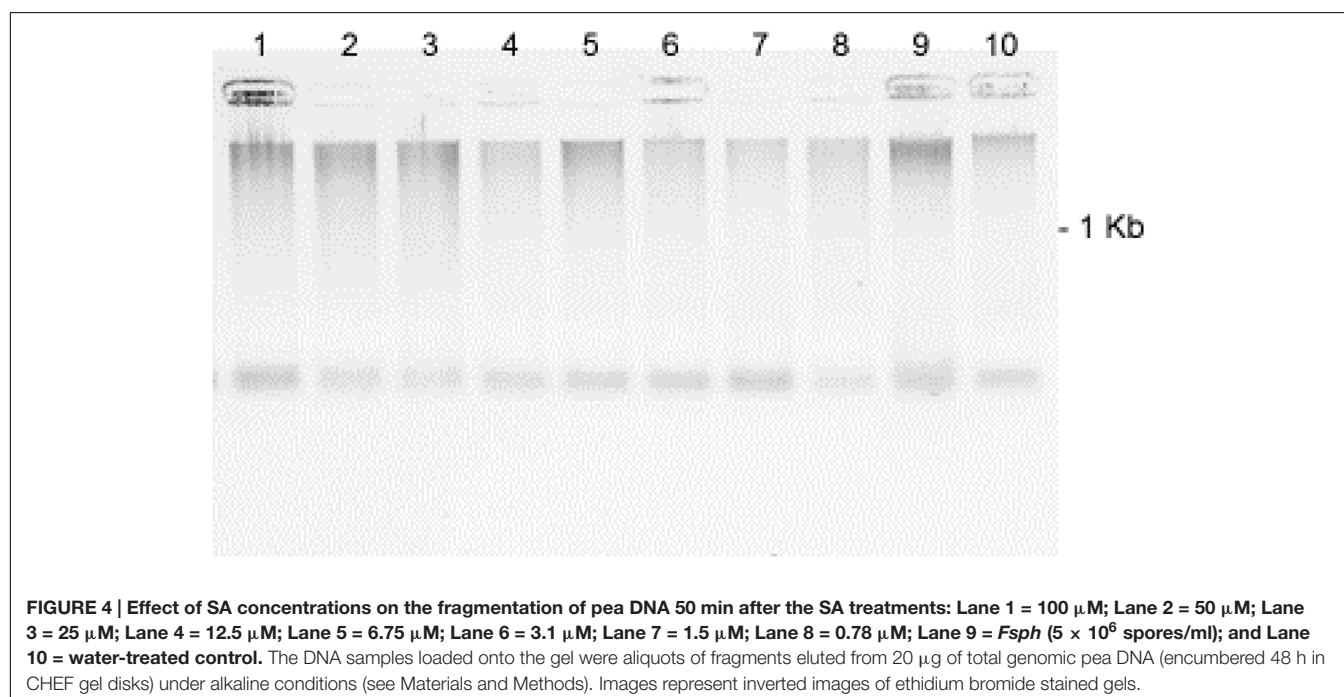


FIGURE 4 | Effect of SA concentrations on the fragmentation of pea DNA 50 min after the SA treatments: Lane 1 = 100 μM; Lane 2 = 50 μM; Lane 3 = 25 μM; Lane 4 = 12.5 μM; Lane 5 = 6.75 μM; Lane 6 = 3.1 μM; Lane 7 = 1.5 μM; Lane 8 = 0.78 μM; Lane 9 = *FspH* (5×10^6 spores/ml); and Lane 10 = water-treated control. The DNA samples loaded onto the gel were aliquots of fragments eluted from 20 μg of total genomic pea DNA (encumbered 48 h in CHEF gel disks) under alkaline conditions (see Materials and Methods). Images represent inverted images of ethidium bromide stained gels.

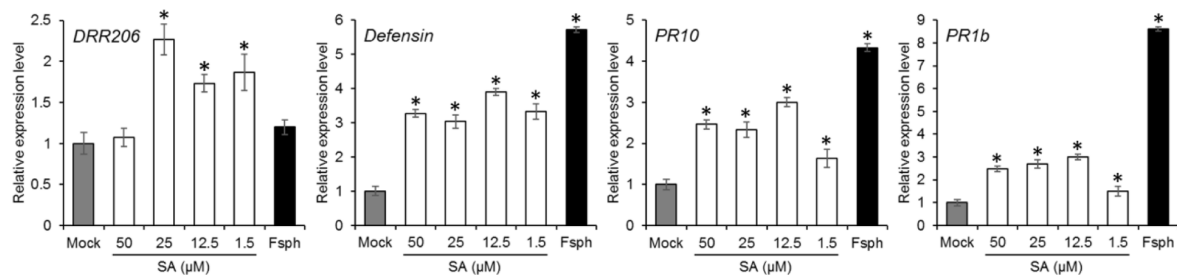


FIGURE 6 | The effect of SA treatment on pea PR gene expression. Pea endocarp tissues were treated for 50 min with water (Mock); 50, 25, 12.5, or 1.5 μ M of SA; or 5×10^6 spores/ml of fungal spores (*FspH*). The tissues were then subjected to qRT-PCR analysis in order to measure the transcriptional responses of the pea PR genes. The data were normalized by the reference gene ubiquitin and converted to a value relative to that of the mock treatment. Histograms represent the means with SE in three replicated experiments.

TABLE 3 | The effect of a high concentration range of SA on the production of pisatin in pea endocarps.

Treatment ^a	Concentration	Pisatin ^b (μ g/g fw)
Water	–	0.0 \pm 0.0
SA	1000 μ M	7.7 \pm 0.02
SA	500 μ M	5.2 \pm 0.59
SA	250 μ M	6.6 \pm 2.5
SA	125 μ M	5.8 \pm 2.3
SA	62 μ M	2.6 \pm 2.6
SA	31 μ M	4.9 \pm 0.2
SA	15 μ M	13.6 \pm 4.5
<i>FspH</i>	1×10^6 spores/ml	113.1 \pm 20.2

^aTreatments (25 μ l) were applied to pea pod halves (~250 mg fresh weight) with the indicated concentrations and subsequently distributed on the surface with a glass rod. Pods were retained in high humidity for 24 h.

^bPisatin was extracted from pea tissue with hexanes. The hexanes were removed by volatilization, and the pisatin-containing residue was extracted with 95% ethanol and quantified at 309 nm.

TABLE 4 | Effect of a lower concentration range of SA on the 24 h production of pisatin in pea endocarp tissue.

Treatment	Concentration	Pisatin (μ g/g fw)
SA	100 μ M	3.1 \pm 0.8
SA	50 μ M	2.0 \pm 1.4
SA	25 μ M	2.6 \pm 0.6
SA	12.5 μ M	3.3 \pm 0.5
SA	6.2 μ M	4.1 \pm 1.3
SA	3.1 μ M	7.5 \pm 1.1
SA	1.5 μ M	19.3 \pm 4.4
SA	0.7 μ M	8.3 \pm 3.4
<i>FspH</i>	2.4×10^6 spores/ml	145 \pm 3.0

Legend is the same as that of Table 3.

DNA in plant cells. Single-strand nicks may have previously escaped observance due to their low abundance within the total, large genomic DNA yields from plants. The alkaline processing and agarose trapping of the total DNA enabled the detection of released DNA fragments (see Materials and Methods) that occur in the very early hours of the inductive treatments

TABLE 5 | Assessment of SA additivity to the synthesis of *FspH*-induced pisatin in pea endocarp tissue after 24 h.

Treatment and molarity ^a	Pisatin (μ g/g fw)
Water	0.0
Water + <i>FspH</i> spores	121.3
SA 100 μ M	2.3
SA 100 μ M + <i>FspH</i> spores	198.8
SA 6.7 μ M	5.9
SA 6.7 μ M + <i>FspH</i> spores	195.2
SA 3.1 μ M	0.7
SA 3.1 μ M + <i>FspH</i> spores	174.8
SA 1.5 μ M	0.0
SA 1.5 μ M + <i>FspH</i> spores	206.0
SA 0.7 μ M	0.02
SA 0.7 μ M + <i>FspH</i> spores	218.3

^aThe indicated treatments were applied (25 μ l) to the endocarp layer of each pea pod half with, when indicated, 5 μ l of *FspH* spores (1×10^7 /ml). Pisatin was extracted (at 24 h) in hexane. The hexane free residue was dissolved in 95% ethanol and quantified at UV₃₀₉.

(Figure 5). This fragmentation was temporally associated with the initiation of defense responses. The sheer presence of the SA association with host DNA is not likely to result in DNA alterations without the presence of a contributing factor such as the direct *in vitro* association of various SA concentrations.

Mechanisms for Regulating PR Gene Expression

Pea PR genes map to multiple chromosomes and often reside in regions that also map as QTLs (Pilet-Nayel et al., 2002). PR genes are ubiquitously present in plant genomes and possess antifungal properties, and therefore, they are potentially the major contributors (Chiang and Hadwiger, 1991; Almeida et al., 2000) to disease resistance. It appears that it is the additive effect of multiple PR genes that develops the complete non-host resistance. Additionally, the pea PR genes analyzed share some homology with the PR genes induced by SA in *Arabidopsis* (Sels et al., 2008).

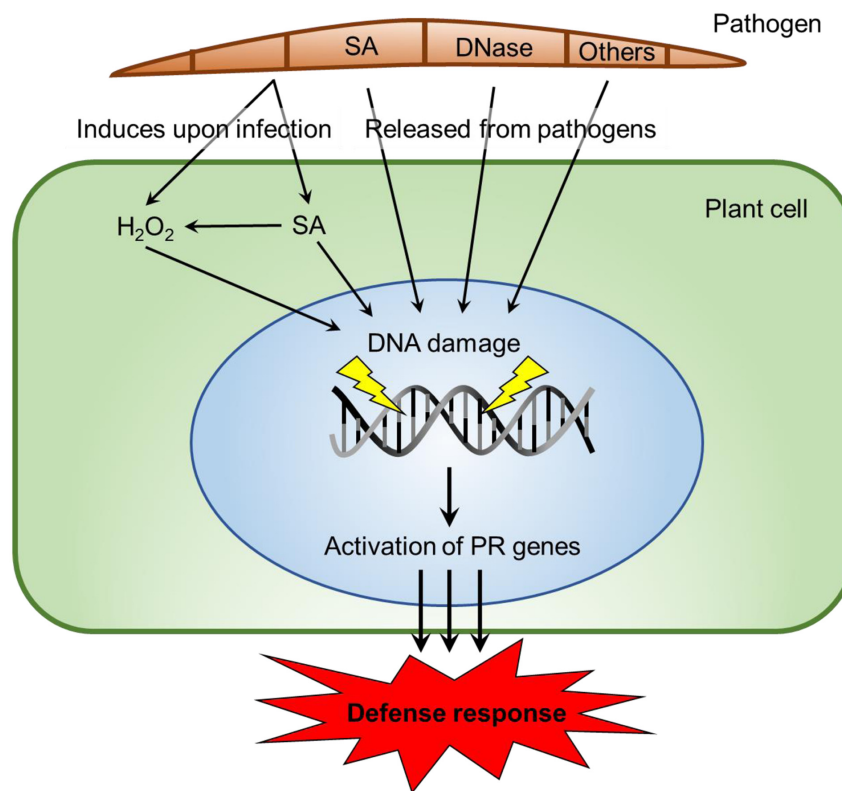


FIGURE 7 | Possible origins of SA signals and mechanisms of DNA damage in host-parasite interactions. SA is produced by demethylation of Methyl-SA or by biosynthesis from cinamate or isochorismate in plants (Raskin, 1992). Alternatively, SA is produced via chorismate in bacteria and fungi (Harper and Hamilton, 1988). Fungi have DNases released (Hadwiger and Polashock, 2013) or other biotic agents. Those agents elicit DNA damage or conformational changes (Hartney et al., 2007), via ROS production at one stage (Driessens et al., 2009; Tanaka and Hadwiger, 2017). We propose that DNA-damaging elicitors can act singly or with assistance at differing DNA sites and may potentially activate PR genes. The elicitor insults to chromatin structure have the ability to restart stalled RNA polymerase complexes.

Role of Chromatin

Gene expression is initiated within chromatin, the site of transcription. The DNA transcription within the region of defense genes can be up regulated or down regulated depending on the associated chromatin structure. Chromatin is a complex of proteins and DNA packed into nucleosomes (Li et al., 2007). The DNA of a particular gene can be accessed by transcription complexes following the alteration of DNA supercoiling or modifications of the relevant proteins. In pea endocarp tissue, PR gene activation is influenced by DNA alterations, by ubiquitination/histone modifications and by a reduction in the architectural transcription factor HMG A (Klosterman et al., 2003; Isaac et al., 2009).

DNA damage can result in the stalling of elongating RNA polymerase II (Lagerwerf et al., 2011) and the attraction of chromatin remodelers to damaged sites. Chromatin modifications function by either disrupting chromatin contacts or affecting the recruitment of non-histone proteins to chromatin (Kouzarides, 2007). Histone modifications can dictate the higher-order chromatin structure in which DNA is packaged, affecting many biological processes. Chromatin structure itself imposes obstacles on all aspects of transcription that are mediated by

RNA polymerase II (Li et al., 2007). The resultant chromatin regulation affects the binding of transcription factors and the initiation and elongation steps of transcription. SA reportedly has an affinity for DNA, and similar to other previously described DNA-specific agents, it can cause DNA damage (Neauault et al., 1996; Yan et al., 2013). The somewhat preferential transcription of PR genes gives credence to the observed selective expression of plant defense resulting from general challenges to sensitive chromatin structures. For transcription to occur, the DNA helix must be opened as the polymerase threads the separated strands through the enzyme (Ma et al., 2013). The upstream torque disrupts the DNA double strand structure and stalls the polymerase, while release of this torsional stress allows the polymerase to resume transcription.

Other DNA Specific Elicitors

Chitosan, a fungal-derived elicitor of PR genes, can compete with histones for sites on DNA and can insert itself into the minor groove of DNA. Fungal DNase (*Fsph* DNase), a second major elicitor of PR gene expression, causes single strand cleavage in double stranded DNA, enabling the release of tension within the DNA helical structures (Gerhold et al., 1993).

Multiple regulatory substances are released from fungal spores following inoculation on their respective host tissues. Of current interest are the proteins that exit the fungal cell via their SignalP sequences (Ramachandran et al., 2016). The functional properties of defense proteins (described above) can range from those that specialize in digesting the cell wall barriers to those that are metabolic enzymes or proteins with unknown function. The regulatory functions affected within the host tissue may also occur either through receptors and subsequent signal cascades that target transcription factors or as direct insults to the organization of sites within chromatin that result in increases in transcription of host genes (Hartney et al., 2007). Although the signaling of compounds of fungal origin can be shown to specifically complex with membrane proteins and modulate the plant's response, the resulting signaling cascade to the site of defense gene transcription is less well understood than in many other host/parasite interactions (Presti et al., 2015). High throughput genomic analyses may be able to detect a multiplicity of potential effectors with the potential to traverse the host-parasite barrier. If so, future research should focus on determining which effectors display potential to play a major role in the processes that result in the development of resistance or enable a susceptible response.

The biotic and abiotic elicitors of PR genes, such as the single strand cleaving DNase elicitor from *FspH*, require a SignalP sequence (Hadwiger and Polashock, 2013). Since homologous fungal genes for DNase production are present in all of the fungal genomes sequenced to date, this chromatin modeling may be implicated in similar signaling in many other plant/fungal interactions (Hadwiger and Polashock, 2013). DNase activity has also been shown to be released from spores of rust (*Puccinia striiformis*), *Verticillium dahliae*, *Colletotrichum coccodes* and yeast cells (Hadwiger and Polashock, 2013). The universality of this enzyme suggests that it could be a general elicitor of the non-host resistance response, protecting plants from pathogens known to be out of their host range. DNase enzymatic action has also been implicated in the initiation of plant defense gene transcription by directly altering nuclear chromatin via single DNA strand nicking. The resultant DNA damage must be subtle enough to alter the chromatin structure for the benefit of the pathogen but not initiate processes that could cause immediate cell death.

CONCLUSION

Salicylic acid is a signal that induces a defense response in *Arabidopsis* and some other plant species (Heil and Bostock,

2002). The data presented also indicate that SA can activate a defense response in pea that is associated with the activation of pea PR genes possessing partial homology with those in *Arabidopsis*. Similar to the signals that activate genes in pea, there is a surge in ROS release within 40 min and temporal DNA damage within 90 min that is detectable in pea DNA fragmentation, in addition to changes in its nuclear appearance and diameter. Although the phytoalexin accumulation is only slightly affected by SA, the effects on the transcription of pea PR genes via DNA damage and distortion may indicate that a signaling route targeting host DNA implicates a different type of chromatin remodeling or transcription initiation. SA may also complement the transcriptional enhancing effect directly on DNA by utilizing a membrane receptor and a subsequent cascade of events that alter the transcription complex by transcription factor attraction or modification. This report shows that ROS that are capable of DNA modification are released. There were nuclear and DNA alterations similar to the changes in other systems that have been associated with enhanced transcription. Additionally, these changes are temporal in the phase that is crucial for the activation of PR genes and non-host resistance in pea endocarp tissue.

AUTHOR CONTRIBUTIONS

LH conceived and designed the experiments, LH and KT conducted the experiment, data analysis, presentation, and wrote the manuscript.

FUNDING

This work was partly supported by Biologically-Intensive Agriculture and Organic Farming (BIOAg) grant from the Center for Sustaining Agriculture and Natural Resources (CSANR) at Washington State University. PPNS No. 0734, Department of Plant Pathology, College of Agricultural, Human, and Natural Resource Sciences, Agricultural Research Center, Hatch Project No. WNP03847, Washington State University, Pullman, WA 99164-6430, USA.

ACKNOWLEDGMENT

Special thanks to Lyndon Porter for the *Fusarium* isolate and Mike Adams and Natalia Moroz for reviewing the manuscript.

REFERENCES

- Almeida, M. S., Cabral, K. M. S., Zingali, R. B., and Kurtenbach, E. (2000). Characterization of two novel defense peptides from pea (*Pisum sativum*) seeds. *Arch. Biochem. Biophys.* 378, 278–286. doi: 10.1006/abbi.2000.1824
- Bailey, J. A., and Mansfield, J. W. (1982). *Phytoalexins*. New York, NY: John Wiley and Sons, 133–180.
- Bau, J. T., Kang, Z., Austihr, C. A., and Kurz, E. U. (2013). SA, a catalytic inhibitor of topoisomerase II inhibits DNA cleavage and is selective for the alpha isoform. *Mol. Pharmacol.* 85, 198–207. doi: 10.1124/mol.113.088963
- Boller, T., and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* 60, 379–406. doi: 10.1146/annurev.arplant.57.032905.105346

- Chang, M. M., Horovitz, D., Culley, D., and Hadwiger, L. A. (1995). Molecular cloning and characterization of a pea chitinase gene expressed in response to wounding, fungal infection and the elicitor chitosan. *Plant Mol. Biol.* 28, 105–111. doi: 10.1007/BF00042042
- Chen, Z., Zheng, Z., Haung, L. A., and Fan, B. (2009). Biosynthesis of salicylic acid in plants. *Plant Signal. Behav.* 4, 493–496. doi: 10.4161/psb.4.6.8392
- Chiang, C. C., and Hadwiger, L. A. (1991). The *Fusarium solani*-induced expression of a pea gene family encoding high cysteine content proteins. *Mol. Plant Microbe Interact.* 4, 324–331. doi: 10.1094/MPMI-4-324
- Choi, J. J., Klosterman, S. J., and Hadwiger, L. A. (2001). A comparison of the effects of DNA-damaging agents and biological elicitors on the induction of plant defense genes, nuclear distortion and cell death. *Plant Physiol.* 125, 752–762. doi: 10.1104/pp.125.2.752
- Coquoz, J. L., Buchala, A., and Matraux, J. B. (1988). The biosynthesis of salicylic acid in potato plants. *Plant Physiol.* 117, 1095–1101. doi: 10.1104/pp.117.3.1095
- De Dieuleveult, M., Yen, K., Hmitou, I., Depaux, A., Boussouar, F., Dargham, C. B., et al. (2016). Genome-wide nucleosome specificity of chromatin remodelers in mouse embryonic stem cells. *Nature* 530, 113–116. doi: 10.1038/nature16505
- Driessens, D., Versteyhe, S., Ghaddhab, C., Burniat, A., De Deken, X., Van Sande, J., et al. (2009). Hydrogen peroxide induces DNA single- double-strand breaks in thyroid cells and is therefore a potential mutagen for this organ. *Endocr. Relat. Cancer* 16, 845–856. doi: 10.1677/erc-09-0020
- Gasser, S., Orsulic, S., Brown, E. J., and Raulet, D. H. (2005). The DNA damage pathway regulates innate immune system ligands of the NKG2D receptor. *Nature* 436, 1186–1190. doi: 10.1038/nature03884
- Gerhold, D. L., Pettinger, A. J., and Hadwiger, L. A. (1993). Characterization of a plant stimulated nuclease from *Fusarium solani*. *Physiol. Mol. Plant Pathol.* 43, 33–46. doi: 10.1006/pmpp.1993.1038
- Grant, J. J., and Loake, G. J. (2000). Role of reactive oxygen intermediates and cognate redox signaling in disease resistance. *Plant Physiol.* 124, 21–29. doi: 10.1104/pp.124.1.21
- Hadwiger, L. A. (2008). Pea-*Fusarium solani* interactions, contributions of a system toward understanding disease resistance. *Phytopathology* 98, 372–379. doi: 10.1094/PHYTO-98-4-0372
- Hadwiger, L. A. (2009). Localization predictions for gene products involved in non-host resistance responses in a model plant/fungal pathogen interaction. *Plant Sci.* 177, 257–265. doi: 10.1016/j.plantsci.2009.06.010
- Hadwiger, L. A. (2015a). Anatomy of a nonhost disease resistance response of pea to *Fusarium solani*: PR gene elicitation via DNase, chitosan and chromatin alterations. *Front. Plant Sci.* 6:373–394. doi: 10.3389/fpls.2015.00373
- Hadwiger, L. A. (2015b). Nonhost resistance: self-inflicted DNA damage by fungal DNase accumulation is a major factor in terminating fungal growth in the pea-*Fusarium solani* f.sp. *phaseoli* interaction. *Physiol. Mol. Plant Pathol.* 92, 79–87. doi: 10.1016/j.pmpp.2015.08.003
- Hadwiger, L. A., and Adams, M. J. (1978). Nuclear changes associated with the host-parasite interaction between *Fusarium solani* and peas. *Physiol. Plant Pathol.* 12, 63–72. doi: 10.1016/0048-4059(78)90019-X
- Hadwiger, L. A., and Chang, M.-M. (2015). Low levels DNA damage occurs as PAMPs, chitin and flg22, activate PR genes, and increase pisatin and disease resistance in pea endocarp tissue. *New Negat. Plant Sci.* 1–2, 6–15. doi: 10.1016/J.neps.201504.001
- Hadwiger, L. A., and Polashock, J. (2013). Fungal mitochondrial DNases. Effectors with the potential to activate plant defenses in nonhost resistance. *Phytopathology* 103, 81–90. doi: 10.1094/PHYTO-04-12-0085-R
- Hadwiger, L. A., and Tanaka, T. (2015). EDTA a novel inducer of pistin, phytoalexin indicator of the nonhost resistance in peas. *Molecules* 15, 24–34. doi: 10.3390/molecules20010024
- Harper, D. B., and Hamilton, F. T. G. (1988). Biosynthesis of chloromethane in *Phellinus pomaceus*. *J. Gen. Microbiol.* 134, 2831–2839. doi: 10.1099/00221287-134-10-2831
- Hartney, S., Carson, J., and Hadwiger, L. A. (2007). The use of chemical genomics to detect functional systems affecting the non-host disease resistance of pea to *Fusarium solani* f.sp. *phaseoli*. *Plant Sci.* 172, 45–56. doi: 10.1016/j.plantsci.2006.07.014
- Heil, M., and Bostock, R. M. (2002). Induced systemic resistance (ISR) against pathogens in the context of induced plant defenses. *Ann. Bot.* 89, 503–512. doi: 10.1093/aob/mcf076
- Isaac, J., Hartney, S. L., Druffel, K., and Hadwiger, L. A. (2009). The non-host disease resistance response in peas; alterations in phosphorylation and ubiquitination of HMG A and histones H2A/H2B. *Plant Sci.* 177, 439–449. doi: 10.1016/j.plantsci.2009.07.007
- Kendra, D. G., Christian, D., and Hadwiger, L. A. (1989). Chitosan oligomers from *Fusarium solani*/pea interactions, chitinase/ β -glucanase digestion of sporelings and from fungal wall chitin actively inhibit fungal growth and enhance disease resistance. *Physiol. Mol. Plant Pathol.* 35, 215–230. doi: 10.1016/0885-5765(89)90052-0
- Klosterman, S. J., Chen, J., Choi, J. J., Chen, E. E., and Hadwiger, L. A. (2001). Characterization of a 20 kDa DNase elicitor from *Fusarium solani* f.sp. *phaseoli* and its expression at the onset of induced resistance in *Pisum sativum*. *Mol. Plant Pathol.* 2, 147–158. doi: 10.1046/j.1364-3703.2001.00062.x
- Klosterman, S. J., Choi, J. J., and Hadwiger, L. A. (2003). Analysis of pea HMG-I/Y (HMG A) expression suggests a role in defense gene regulation. *Mol. Plant Pathol.* 4, 249–258. doi: 10.1046/J.1364-3703.2003.00171.X
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705. doi: 10.1016/j.cell.2007.02.005
- Lagerwerf, S., Vrouwe, M. G., Overmeer, R. M., Fouteri, M. I., and Mullenders, L. H. F. (2011). DNA damage response and transcription. *DNA Repair* 10, 743–750. doi: 10.1016/j.dnarep.2011.04.024
- Li, B., Carey, M., and Workman, J. L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719. doi: 10.1016/j.cell.2007.01.015
- Li, X., Juang, Y., Ji, Z., Liu, Y., and Zhang, Q. (2015). BRHIS1 suppresses rice innate immunity through binding to monoubiquitinated H2A and H2B variants. *EMBO Rep.* 16, 1192–1202. doi: 10.15252/embr.201440000/
- Ma, J., Bai, L., and Wang, M. D. (2013). Transcription under torsion. *Science* 340, 1580–1583. doi: 10.1126/science.1235441
- Maeda, H., and Dudareva, N. (2012). The shikimate pathway and aromatic amino acid biosynthesis in plants. *Ann. Rev. Plant Biol.* 63, 73–105. doi: 10.1146/annurev-arplant-042811-105439
- McDowell, J. M., and Woffenden, B. J. (2003). Plant disease resistance genes: recent insights and potential applications. *Trends Biotechnol.* 21, 178–183. doi: 10.1016/S0167-7799(03)00053-2
- Neault, J. F., Naoui, M., Manfait, M., and Tajmir-Riahi, H. A. (1996). Aspirin-DNA interaction studied by FTIR and laser Raman difference spectroscopy. *FEBS Lett.* 382, 26–30. doi: 10.1016/0014-5793(96)00093-2
- Orozco-Cardenas, M., and Ryan, C. A. (1999). Hydrogen peroxide is generated systemically in plant leaves by wounding and systemin via the octadecanoid pathway. *Proc. Natl. Acad. Sci. U.S.A.* 96, 6553–6557. doi: 10.1073/pnas.96.116553
- Pilet-Nayel, M. L., Meuhlbaier, F. J., McGee, R. J., Kraft, J. M., Baranger, A., and Coyne, C. J. (2002). Quantitative trait loci for partial resistance to *Aphanomyces* root rot in pea. *Theor. Appl. Genet.* 106, 28–39. doi: 10.1007/s00122-002-0985-2
- Presti, L. L., Lanver, D., Schweizer, G., Tanaka, S., Liang, L., Tollot, M., et al. (2015). Fungal effectors and Plant susceptibility. *Annu. Rev. Plant Biol.* 66, 513–543. doi: 10.1146/annurev-arplant-043014-114623
- Ramachandran, S. R., Yen, C., Kud, J., Tanaka, K., Mahoney, A. K., Xiao, F., et al. (2016). Effectors from wheat rust fungi suppress multiple plant defense responses. *Phytopathology* 107, 75–83. doi: 10.1094/PHYTO-02-16-0083-R
- Raskin, I. (1992). Role of salicylic acid in plants. *Annu. Rev. Plant Physiol.* 43, 438–463. doi: 10.1146/annurev.pp.43.060192.002255
- Sels, J., Mathys, J., De Coninck, B. M. A., Cammue, B. P. A., and De Bolle, M. F. C. (2008). Plant pathogenesis-related (PR) proteins: A focus on PF peptides. *Plant Physiol. Biochem.* 46, 941–950. doi: 10.1016/j.plaphy.2008.06.011
- Song, J., and Bent, A. F. (2014). Microbial pathogens trigger host DNA double-strand breaks whose abundance is reduced by plant defense responses. *PLoS Pathog.* 10:e1004030. doi: 10.1371/journal.ppat.1004030
- Stewart, G. S., Panier, S., Townsend, K., Al-Hakim, A. K., Kolas, N. K., Miller, E. S., et al. (2009). The RIDDLE syndrome protein mediates a ubiquitin-dependent signaling cascade at sites of DNA damage. *Cell* 136, 420–434. doi: 10.1016/j.cell.2008.12.042
- Tanaka, K., and Hadwiger, L. A. (2017). Nonhost resistance: reactive oxygen species (ROS) signal causes DNA damage prior to the induction of PR genes and disease

- resistance in pea tissue. *Physiol. Mol. Plant Pathol.* 98, 18–24. doi: 10.1016/j.pmpp.2017.01.007
- van Loon, L. C. (1985). Pathogenesis-related proteins. *Plant Mol. Biol.* 4, 111–116. doi: 10.1007/BF02418757
- Wu, Y., Zhang, D., Chu, J. Y., Boyle, P., Wang, Y., Brindle, I. D., et al. (2012). The Arabidopsis NPR1 protein is a receptor for the plant defense hormone salicylic acid. *Cell Rep.* 1, 639–647. doi: 10.1016/j.celrep.2012.05.008
- Yan, S., Wang, W., Marques, J., Mohan, R., Saleh, A., Durrant, W. E., et al. (2013). Salicylic acid activates DNA damage responses to potentiate plant immunity. *Mol. Cell* 52, 602–610. doi: 10.1016/j.molcel.2013.09.019
- Zhou, J., Tang, X., and Martin, G. B. (1997). The Pto kinase conferring resistance to tomato bacterial speck disease interacts with proteins that bind a cis-element of pathogenesis-related genes. *EMBO J.* 16, 3207–3218. doi: 10.1093/emboj/16.11.3207
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hadwiger and Tanaka. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genome-Wide Dissection of the Heat Shock Transcription Factor Family Genes in *Arachis*

Pengfei Wang¹, Hui Song¹, Changsheng Li¹, Pengcheng Li¹, Aiqin Li¹, Hongshan Guan¹, Lei Hou^{1*} and Xingjun Wang^{1,2*}

¹ Biotechnology Research Center, Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan, China, ² College of Life Sciences, Shandong Normal University, Jinan, China

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Benedetto Ruperti,
University of Padova, Italy
Hui Wang,
University of Georgia, USA

*Correspondence:

Lei Hou
houlei9042@163.com
Xingjun Wang
xingjunw@hotmail.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 18 September 2016

Accepted: 18 January 2017

Published: 06 February 2017

Citation:

Wang P, Song H, Li C, Li P, Li A,
Guan H, Hou L and Wang X (2017)
Genome-Wide Dissection of the Heat
Shock Transcription Factor Family
Genes in *Arachis*.
Front. Plant Sci. 8:106.
doi: 10.3389/fpls.2017.00106

Heat shock transcription factors (Hsfs) are important transcription factors (TFs) in protecting plants from damages caused by various stresses. The released whole genome sequences of wild peanuts make it possible for genome-wide analysis of Hsfs in peanut. In this study, a total of 16 and 17 *Hsf* genes were identified from *Arachis duranensis* and *A. ipaensis*, respectively. We identified 16 orthologous *Hsf* gene pairs in both peanut species; however *HsfXs* was only identified from *A. ipaensis*. Orthologous pairs between two wild peanut species were highly syntenic. Based on phylogenetic relationship, peanut Hsfs were divided into groups A, B, and C. Selection pressure analysis showed that group B *Hsf* genes mainly underwent positive selection and group A Hsfs were affected by purifying selection. Small scale segmental and tandem duplication may play important roles in the evolution of these genes. Cis-elements, such as ABRE, DRE, and HSE, were found in the promoters of most *Arachis* *Hsf* genes. Five *AdHsfs* and two *AiHsfs* contained fungal elicitor responsive elements suggesting their involvement in response to fungi infection. These genes were differentially expressed in cultivated peanut under abiotic stress and *Aspergillus flavus* infection. *AhHsf2* and *AhHsf14* were significantly up-regulated after inoculation with *A. flavus* suggesting their possible role in fungal resistance.

Keywords: heat shock transcription factor, peanut, abiotic stress, selective pressure, purifying selection

INTRODUCTION

Abiotic stresses, including heat, cold, drought, and salinity, affect plant growth and development and cause serious loss of crop production (Wang et al., 2004; Al-Whaibi, 2011; Qiao et al., 2015). As sessile organisms, plants could not change their locations when facing such stress conditions (Guo et al., 2016). However, plants have evolved adaptation strategies to these stresses (Scharf et al., 2012; Guo et al., 2016). Transcription factors play a crucial role in stress tolerance by regulating the expression of thousands of genes under unfavorable conditions (Schwechheimer and Bevan, 1998; Kreps et al., 2002; Shinozaki and Yamaguchi-Shinozaki, 2007; Wang et al., 2014). Plant heat shock transcription factors (Hsfs) are important transcription factors (TFs) in protecting plants from heat stress and other stresses, including cold, salinity, and drought (Kotak et al., 2007a; Swindell et al., 2007; Hu et al., 2015). Hsfs were found in eukaryotes from yeast to humans (Ritossa, 1962; Tanabe et al., 1997; Akerfelt et al., 2010). Hsfs could protect cells from extreme proteotoxic damage via the

activation of related genes (Dalton et al., 2000; Akerfelt et al., 2010; Yang et al., 2014; Jaeger et al., 2016). Studies showed that Hsfs are also involved in plant growth and development (Almoguera et al., 2002; Díaz-Martín et al., 2005; Kotak et al., 2007b).

Hsfs regulate heat shock response via activating the expression of heat shock protein (HSP) genes by binding to the heat shock elements (HSEs) (Pelham, 1982; Akerfelt et al., 2010). The sequences and geometries of HSEs (5'-AGAAnnTTCT-3') are variable (Guertin et al., 2012; Mendillo et al., 2012; Vihervaara et al., 2013). Hsfs could also bind to SatIII repeat element, 5'-cgGAATgGAATg-3' (Grady et al., 1992). Like many other transcription factors, Hsfs have an N-terminal DNA binding domain (DBD) and followed by an oligomerization domain (OD). OD is composed of two hydrophobic heptad repeats (HR-A/B) which allows homo- and hetero-multimerization (Peteranderl et al., 1999; Nover et al., 2001; Baniwal et al., 2004). Certain Hsfs contained nuclear location signal (NLS) domain, nuclear export signal (NES), and C-terminal activation (AHA) domain (Döring et al., 2000; Hsu et al., 2003; Maere et al., 2005). Based on the structural characteristics of HR-A/B domain and the phylogenetic relationship, plant Hsfs are divided into A, B, and C groups (Von Koskull-Döring et al., 2007; Wang et al., 2014; Yang et al., 2014). Additional sequences were found in HR-A/B domain of group A and C, but not in group B Hsfs (Nover et al., 2001; Schmidt et al., 2012; Wang et al., 2014; Yang et al., 2014).

Only the active Hsfs are capable of recognizing and binding to the promoters of target genes. The inactive monomer could be converted into active oligomer under variety of stress conditions (Hartl and Hayer-Hartl, 2002; Wang et al., 2012; Li et al., 2014). There are only a few Hsf genes in yeast and animals, while 20–50 Hsf genes were found in plants (Scharf et al., 2012; Lin et al., 2014; Qiao et al., 2015). Hsf genes were identified in many plants and expressed in various tissues at different developmental stages during different stress conditions (Giorno et al., 2012; Chung et al., 2013; Xue et al., 2014).

Peanut (*Arachis hypogaea* L.) is an important oil crop in the world. In developing countries, peanuts were rain-fed, so it is important to study the drought stress tolerance of peanut (Ramu et al., 2015). *Aspergillus flavus* produces potent mycotoxins known as aflatoxins that could cause serious health concerns (Zhang et al., 2015). It is unknown on the role of Hsf genes in peanut response to abiotic stresses and *A. flavus* infection.

Cultivated peanut is an allotetraploid (AABB, $4n = 4x = 40$) originated from a single hybridization and genome duplication event between two wild type diploid peanuts (AA and BB genomes) (Kochert et al., 1996; Freitas et al., 2007; Moretzsohn et al., 2013; Wang et al., 2016). Recently, the whole genome sequencing of the two ancestral species (*A. duranensis* and *A. ipaensis*) have been completed (Bertioli et al., 2016; <http://peanutbase.org/>). Here, we genome-widely identified and analyzed the *Hsf* genes from two wild peanuts species: *A. duranensis* (AA genome) and *A. ipaensis* (BB genome), respectively. We analyzed the gene duplication events in the wild peanut species, the difference of selection pressure in A, B, and C group of *Arachis* Hsfs, and the structures of these proteins. Our results provide basic information for further understanding

the functional divergence and evolution of *Arachis* Hsfs. We also applied the knowledge gained from wild species to cultivated one to understand their possible functions on peanut response to abiotic and biotic stress.

MATERIALS AND METHODS

Data Collection and Identification of *Hsf* Genes

The genome sequence data of two wild peanut species (AA and BB genomes) were obtained from the peanut genome database (<http://peanutbase.org/>). The conserved domains of Hsfs are Hsf-type DBD domain. The HMM ID of this domain is PF00447 in the pfam database (<http://pfam.xfam.org/>). The amino acid sequences of HMMs were used as queries to identify all possible Hsf protein sequences in AA and BB genome database using BLASTP ($E < 0.001$). SMART software (<http://smart.embl-heidelberg.de/>) was used to identify integrated DBD domain and (HR-A/B) domain in the putative peanut Hsfs. Candidate proteins without integrated DBD domain and HR-A/B domain were removed. NLS domains in peanut Hsfs were predicted using cNLS Mapper software (http://nls-mapper.iab.keio.ac.jp/cgi-bin/NLS_Mapper_form.cgi). NES domains were predicted using NetNES 1.1 server software (<http://www.cbs.dtu.dk/services/NetNES/>). AHA domains were predicted based on the conserved-type AHA motif sequence FWxxF/L, F/I/L (Kotak et al., 2004). Protein isoelectric point (pI) and molecular weight (Mw) were analyzed using ExPASy software (http://web.expasy.org/compute_pi/).

The genome, protein, and cDNA sequences were collected from the related genome databases for the following additional plant species: *Arabidopsis thaliana* (<http://www.plantgdb.org/AtGDB/>), *Glycine max* (<http://www.plantgdb.org/GmGDB/>), *Lotus japonicus* (<http://www.plantgdb.org/LjGDB/>), *Medicago truncatula* (<http://www.plantgdb.org/MtGDB/>), *Cajanus cajan* (<http://gigadb.org/dataset/100028>) and *Cicer arietinum* (<http://nigr.res.in/CGAP/home.php>).

Orthologous Gene Identification and Structure Analysis

Orthologous gene pairs were identified according to (1) the best-hit between *A. duranensis* and *A. ipaensis*, (2) the position in the phylogenetic tree (bootstrap value > 50), and (3) identity between orthologous gene pairs ($> 90\%$). Circos software was used to plot the chromosomal location (Krzywinski et al., 2009). Gene Structure Display Server 2.0 (<http://gsds.cbi.pku.edu.cn/>) was used to plot the gene structure.

Analysis of Synteny

Intraspecies synteny analysis of AA or BB genome and interspecies synteny analysis between AA and BB genomes were based on comparison of 100 kb blocks of chromosome containing Hsf genes according to previous reports (Sato et al., 2008; Zhang et al., 2011; Lin et al., 2014). Hsf genes were set as anchor points according to their chromosome locations. Blocks were identified by local all-vs-all BLASTN ($E < 10^{-20}$). In intraspecies analysis, when four or more homology genes were detected,

these two blocks were considered to be originated from a large-scale duplication event (Zhang et al., 2011; Lin et al., 2014). In interspecies analysis, when three or more conserved homology genes were detected, these two blocks were considered syntenic blocks (Sato et al., 2008; Lin et al., 2014; Wang et al., 2016).

Multiple Sequence Alignment and Phylogenetic Analysis

Protein multiple sequence alignment was performed using online software Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo>). Neighbor-Joining (NJ) trees were constructed using MEGA 6.0 with protein sequences. To support the calculated relationship, 1000 bootstrap samples were generated. A total of 21 *A. thaliana* Hsfs (Scharf et al., 2012), 11 *M. truncatula* Hsfs, 10 *L. japonicus* Hsfs, 16 *C. cajan* Hsfs (Lin et al., 2014), 11 *C. arietinum* Hsfs, and 40 *G. max* Hsfs were included in the phylogenetic analysis (Poisson correction, pairwise deletion, and bootstrap = 1000 replicates; Xue et al., 2014). All Hsfs used in this study was listed in the **Table S1**.

Gene Duplication Analysis

Two standards for duplication gene identification were used. High-stringency standard: coding protein pair with $\geq 50\%$ identity and covering $\geq 90\%$ protein length. Low-stringency standard: protein pair with $\geq 30\%$ identity and covering $\geq 70\%$ protein length (Rizzon et al., 2006). Tandem duplication of genes was marked according to the previously described method (Yuan et al., 2015). Chromosome segmental or large scale duplication of genes was identified based on the intraspecies synteny (Zhang et al., 2011; Lin et al., 2014; Qiao et al., 2015).

Protein Structure Analysis and Homology Modeling

SWISS-MODEL (<http://www.swissmodel.expasy.org/interactive>) was used to calculate secondary structure and build three-dimensional structure of proteins. The templates for building protein 3D model were selected in PDB database based on the best identity. Protein 3D models were selected based on the best global model quality estimation (GMQE). Homology modeling templates included 5d5v.1 (monomer of DBD domain), 5d5v.1 (homo-dimer of DBD domain interacted with SalIII), 5d5u.1 (homo-dimer DBD domain interacted with HSE), 4r0r.1.A (monomer of HR-A/B domain) and 4r0r.1 (homo-trimer of HR-A/B).

Analysis of Selective Pressure

Codeml program under PAML (phylogenetic analysis maximum likelihood) version 4.7 software (Yang, 2007) was used to detect whether the Hsf genes underwent positive selection. In PAML, six site models, M0 (one ratio), M1a (neutral), M2a (positive selection), M3 (discrete), M7 (beta) and M8 (beta and ω) could be applied to selection pressure analysis. Positive selection sites could be identified by the comparison of M0-M3, M1a-M2a, and M7-M8 (Yang et al., 2000).

Analysis of Cis-Acting Regulatory Elements in Promoter

Plantcare software (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) was used to predict cis-acting regulatory elements.

Plant Materials, Stress Treatments, and RNA Isolation

Cultivated peanut cv. Luhua-14 was used in this study. Eleven-day-old peanut seedlings were subjected to drought (removed from wet medium and kept in air on filter paper), cold (4°C) and high temperature (42°C) treatment. Leaf samples were collected at 0, 1, and 6 h after treatment and immediately frozen in liquid nitrogen. Leaf samples without treatment were used as control. Peanut seeds inoculated with *A. flavus* for 3 days were collected and seeds without *A. flavus* inoculation were used as control according to a previous report (Zhang et al., 2015). RNAs were isolated by CTAB method according to a previous method (Wang et al., 2016). For reverse transcription, the first-strand cDNA was synthesized with an oligo (dT) primer using a PrimeScript™ first-strand cDNA synthesis kit (TaKaRa). Three technical replicates were carried out in this study.

Gene Expression Analysis

Quantitative real time PCR (qRT-PCR) was performed using the FastStart Universal SYBR Green Master (ROX) with ABI™ 7500. The qRT-PCR program was set as the following: 95°C for 30 s, followed by 40 cycles of 95°C for 5 s, 60°C for 30 s. Relative gene expression levels were calculated using the $\Delta\Delta\text{CT}$ method. The primers for qRT-PCR were provided in the **Table S2**. *T*-test was used to analyze the significance.

RESULTS

Identification of Hsf Genes in Wild Peanut Species

The amino acid sequences of Hsfs were extracted from AA and BB wild peanut genome database using the BLASTP program. The amino acid sequences of Hsf DBD domains (Pfam: PF00447) were used as queries. From AA and BB genomes, we identified 16 and 17 Hsf genes, respectively. The polypeptide lengths of Hsfs varied from 209 to 656 aa in *A. duranensis* and from 282 to 514 aa in *A. ipaensis*. *A. thaliana* Hsf family were often employed as reference to classify Hsf family in other plant species (Scharf et al., 2012; Li et al., 2014; Wang et al., 2014; Qiao et al., 2015). We employed Hsfs from *A. thaliana* and other species to construct phylogenetic tree together with Hsfs in two wild peanut species. In this study, 21 *A. thaliana* Hsfs, 11 *M. truncatula* Hsfs, 10 *L. japonicus* Hsfs, 16 *C. cajan* Hsfs, 11 *C. arietinum* Hsfs, and 40 *G. max* Hsfs were used for phylogenetic tree construction (**Figure 1**). These Hsfs were divided into A, B, and C groups that was consistent with previous studies (Scharf et al., 2012; Li et al., 2014; Lin et al., 2014; Wang et al., 2014; Qiao et al., 2015). Group A was divided into 10 clusters, group B was divided into five clusters, and group C contained only one cluster.

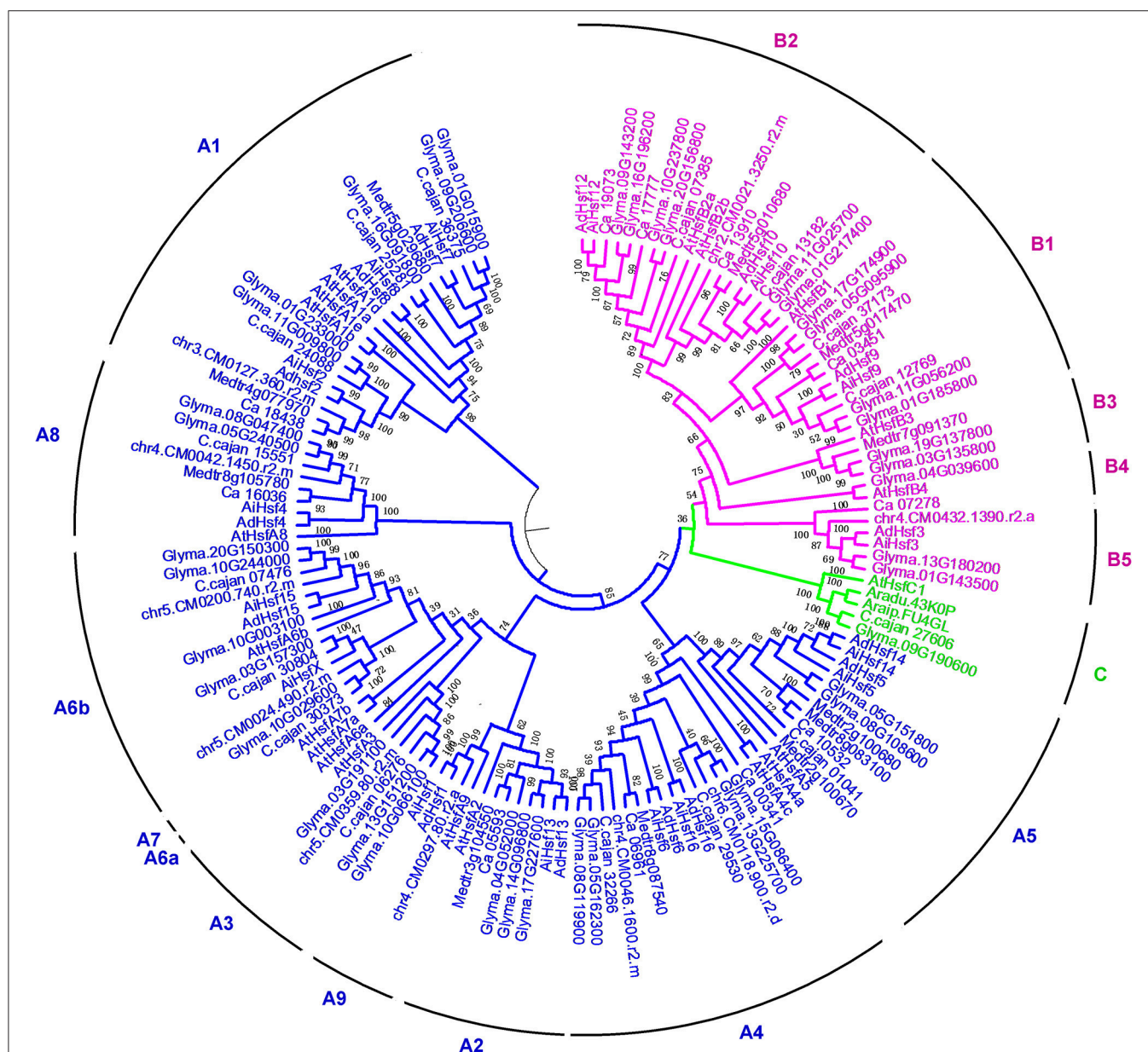


FIGURE 1 | Neighbor-joining phylogenetic analysis of Hsfs.

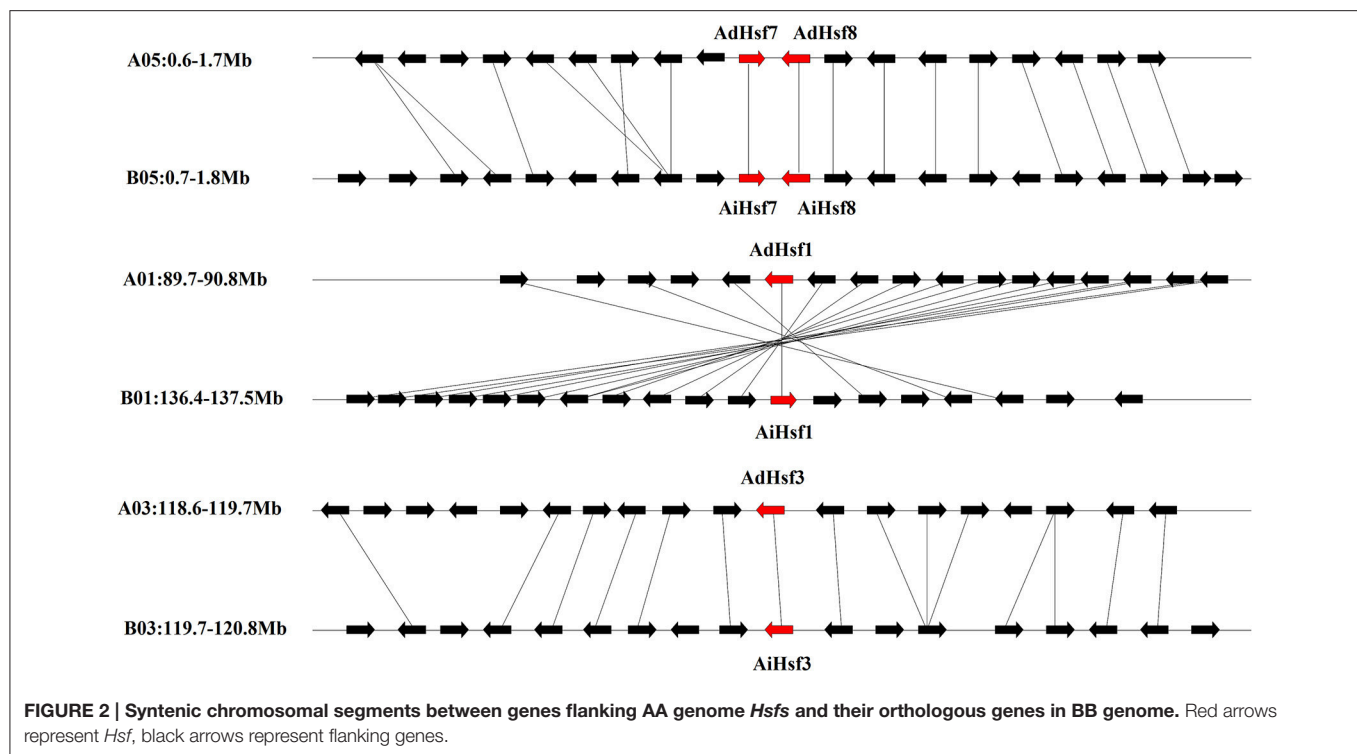
Clusters in the group A were named as A1–A5, A6a, A6b, A7–A9. Clusters in the group B were named as B1–B5. B5 cluster was not presented in *Arabidopsis*; however, B5 cluster was identified in many leguminous species including wild peanut species. In wild peanut species, A3, A6a, A7, B3, and B4 clusters were absent (Figure 1). Orthologous of all 16 AA genome Hsfs were found in the BB genome with >90% identity (Table S3).

Interspecies synteny analysis showed that high level synteny was maintained between AA and BB genomes (Figure 2). This synteny analysis supported the identification of orthologous-pairs of Hsfs between AA and BB genomes. The nomenclature of AA genome Hsfs was based on their chromosome location order, AdHsf1–16. BB genome Hsfs were named based on their

orthologous genes in AA genome AiHsf1–16 and AiHsfX. The orthologous gene of AiHsfX (Araip. A5C77) was not found in AA genome. The gene IDs and physical locations information of wild peanut Hsf genes were showed in Table 1, Figure 3.

Duplication of Hsf Genes in Peanut

Duplicated gene-pairs were found in both AA and BB genomes, including high-stringency standard duplicated gene-pairs AdHsf5–AdHsf14, AdHsf6–AdHsf16 in AA genome and AiHsf5–AiHsf14, AiHsf6–AiHsf16, AiHsf7–AiHsf8 in BB genome, low-stringency standard duplicated gene-pairs AdHsf7–AdHsf8 in AA genome and AiHsf15–AiHsfX in BB genome. Intrasynteny analysis showed that the duplicated gene-pair blocks



were not collinear. No chromosome segmental or large scale duplication gene pairs were identified. *AiHsf7-AiHsf8* and *AdHsf7-AdHsf8* were identified as tandem duplicated gene-pairs.

Features of *Hsfs* in Wild Peanut Species

Most members of *Hsf* gene families in both AA and BB genomes contained one intron and two exons. However, *AdHsf7* contained three exons and *AdHsf14* contained four exons in the AA genome, *AiHsf15* contained three exons, *AiHsf14*, and *AiHsfX* contained four exons in the BB genome. *AdHsf14* contained four exons, while its duplicated gene *AdHsf5* contained only two exons. Intronless *Hsfs* were also found in both AA and BB genomes (Figure S1).

HR-A/B domain is critical for one *Hsf* interacting with other *Hsfs* to form trimer through a helical coiled-coil structure (Scharf et al., 2012; Jaeger et al., 2016; Neudegger et al., 2016). Similar to other plant *Hsfs*, group A *Hsfs* have an insertion between HR-A and HR-B regions in peanut. However, this insertion was not found in the group B *Hsfs*. In *Arachis*, the sequence of group B *Hsf* HR-A/B was not conserved compare with that in group A (Figure 4). The DBD domains were conserved in two wild peanut species. The most conserved motif of DBD domains were “FSSFI/VRQLNT/I” in peanut (Figure S2).

The 3D Structure of *Hsfs* in Wild Peanut Species

The predicted 3D structures of BDB domain of all AA and BB *Hsfs* were similar to that of human *Hsf* BDB (Figure 5A). The predicted 3D structures of HR-A/B domain of AA and BB *Hsfs* were also similar to the human *Hsfs* (Figure 5C). The 3D

structures of BDB domain of peanut orthologous were highly conserved.

When adjacent DBD molecules bound to HSE element, two DBD molecules formed symmetrical protein-protein interaction involving the helix $\alpha 2$. The closest intermolecular contact occurred between the Gly50 residues located at the N-terminal end of the $\alpha 2$ helices in chordate *Hsfs*. Gly50 is conserved and is surrounded by Gln49 and Gln51 in chordate *Hsfs* (Neudegger et al., 2016). In peanut, we predicted that the closest intermolecular contact residues by homologous comparison and 3D model comparison. The results showed that the closest contact residues were not conserved between chordate *Hsfs* and peanut *Hsfs*. For example, in *AdHsf1* and *AdHsf5*, the predicted closest intermolecular contact occurred between the residues His143 (Figure 5A). We also built models that DBD domain of AA and BB wild peanut *Hsfs* bound to SatIII element. The result showed that the predicted dimer structures of DBD-DBD interaction for binding to SatIII element and HSE element were distinct (Figure 5B).

Selective Pressure Analysis of *Hsfs* in Wild Peanut Species

Site models were used to detect whether different groups of *Hsfs* were under different selective pressure in peanuts. Group C *Hsfs* contained only one gene, it could not be analyzed. M0 showed that both *AdHsfs* and *AiHsfs* in group A underwent strong purifying selection ($\omega = 0.31723$ in AA genome and $\omega = 0.40488$ in BB genome; Table S4). Interestingly, in group B, both *AdHsfs* and *AiHsfs* were underwent positive selection ($\omega = 1.69713$ in AA genome and $\omega = 1.95226$ in BB genome). M0

TABLE 1 | Hsfs identified in wild type peanuts.

Gene name	Chromosome location	Gene ID
<i>AdHsf12</i>	Aradu.A05:106,432,571..106,434,237	<i>Aradu.H5REB</i>
<i>AdHsf9</i>	Aradu.A05:7,998,864..8,000,946	<i>Aradu.S4DGV</i>
<i>AdHsf10</i>	Aradu.A05:14,375,242..14,377,500	<i>Aradu.NTQ7W</i>
<i>AdHsf1</i>	Aradu.A01:90,085,291..90,088,131	<i>Aradu.5S8J3</i>
<i>AdHsf2</i>	Aradu.A03:8,559,905..8,562,946	<i>Aradu.N3N49</i>
<i>AdHsf16</i>	Aradu.A10:1,011,979..1,014,302	<i>Aradu.H8ZFN</i>
<i>AdHsf15</i>	Aradu.A09:110,970,626..110,972,271	<i>Aradu.MA5WH</i>
<i>AdHsf6</i>	Aradu.A03:134,564,420..134,565,759	<i>Aradu.RVF1V</i>
<i>AdHsf7</i>	Aradu.A05:1,337,984..1,342,330	<i>Aradu.G1NHG</i>
<i>AdHsf4</i>	Aradu.A03:127,573,853..127,576,100	<i>Aradu.N1DAK</i>
<i>AdHsf5</i>	Aradu.A03:133,622,480..133,625,670	<i>Aradu.R0U9B</i>
<i>AdHsf11</i>	Aradu.A05:91,450,473..91,451,748	<i>Aradu.43K0P</i>
<i>AdHsf14</i>	Aradu.A09:108,105,828..108,110,202	<i>Aradu.RFN6Q</i>
<i>AdHsf8</i>	Aradu.A05:1,343,331..1,346,309	<i>Aradu.95F2Q</i>
<i>AdHsf3</i>	Aradu.A03:119,139,289..119,141,972	<i>Aradu.X3DNX</i>
<i>AdHsf13</i>	Aradu.A06:108,820,015..108,822,104	<i>Aradu.RY508</i>
<i>AiHsf7</i>	Araip.B05:1,259,505..1,263,848	<i>Araip.T98BR</i>
<i>AiHsf12</i>	Araip.B05:146,768,456..146,770,030	<i>Araip.8V58J</i>
<i>AiHsf9</i>	Araip.B05:8,412,092..8,414,222	<i>Araip.90YM8</i>
<i>AiHsf10</i>	Araip.B05:15,070,998..15,073,228	<i>Araip.K80UC</i>
<i>AiHsf1</i>	Araip.B01:137,050,298..137,053,902	<i>Araip.DCD5Q</i>
<i>AiHsf2</i>	Araip.B03:11,777,014..11,780,056	<i>Araip.U7I8R</i>
<i>AiHsf16</i>	Araip.B10:2,967,059..2,969,380	<i>Araip.4A18K</i>
<i>AiHsf15</i>	Araip.B09:145,996,176..145,998,569	<i>Araip.24AK5</i>
<i>AiHsfX</i>	Araip.B06:1,771,852..1,780,911	<i>Araip.A5C77</i>
<i>AiHsf6</i>	Araip.B03:135,626,726..135,628,064	<i>Araip.2D8LN</i>
<i>AiHsf4</i>	Araip.B03:128,239,020..128,240,481	<i>Araip.3P0PJ</i>
<i>AiHsf5</i>	Araip.B03:889,869..893,296	<i>Araip.Z1XB6</i>
<i>AiHsf11</i>	Araip.B05:137,352,977..137,354,281	<i>Araip.FU4GL</i>
<i>AiHsf3</i>	Araip.B03:120,142,696..120,145,308	<i>Araip.UGV6F</i>
<i>AiHsf8</i>	Araip.B05:1,265,312..1,268,368	<i>Araip.SMI4I</i>
<i>AiHsf14</i>	Araip.B09:130,155,536..130,159,602	<i>Araip.G2FC5</i>
<i>AiHsf13</i>	Araip.B06:133,072,149..133,074,221	<i>Araip.B0RQS</i>

These gene ID could be searched on web (http://peanutbase.org/keyword_search).

vs. M3, M1a vs. M2a and M7 vs. M8 comparisons detected 399 positive selection sites in group B *AdHsfs* ($P < 0.05$) and 382 positive selection sites in group B *AiHsfs* ($P < 0.001$; **Table S4**). The identification of these positive selection sites in group B *Hsfs* indicated extensive functional diversity and structural variation (Wang et al., 2016).

Cis-Acting Regulatory Element Analysis of Peanut *Hsf* Promoter

In silico survey of the putative cis-acting regulatory elements in the 1500 kb promoter region of *Hsfs* was performed. The majority of *Hsf* promoters contained HSE elements. HSE was not found in *AdHsf3*, *AdHsf12*, *AdHsf14*, *AiHsf1*, or *AdHsf11* promoters. Many *Hsf* promoters except *AiHsf2* contained abiotic stress responsive element such as MBS (drought inducible), LTR (low temperature responsive), and ARE elements (anaerobic

induction). RNA-seq data showed that two *A. duranensis* *Hsf* genes (*Aradu.X3DNX*, *AdHsf3*, and *Aradu.5S8J3*, *AdHsf1*) were up-regulated significantly under drought stress ($\log_2 FC > 2$, $FDR < 0.05$) (Guimarães et al., 2012; Brasileiro et al., 2015). Phytohormone-induced elements, such as ERE element (ethylene-responsive element), AuxRR-core or TGA-element (auxin responsive), GARE-motif, or P-box element (gibberellin-responsive), ABRE element (ABA responsive), TCA-element (salicylic acid responsive), and TGACG-motif or CGTCA-motif element (MeJA-responsive) were found in some *Hsf* promoters. Five *AdHsfs* (*AdHsf2*, *AdHsf4*, *AdHsf16*, *AdHsf8*, and *AdHsf6*) and two *AiHsfs* (*AdHsf11* and *AdHsf10*) contained fungal elicitor responsive elements. Promoters of orthologous genes between AA and BB genomes were similar (**Table S5**).

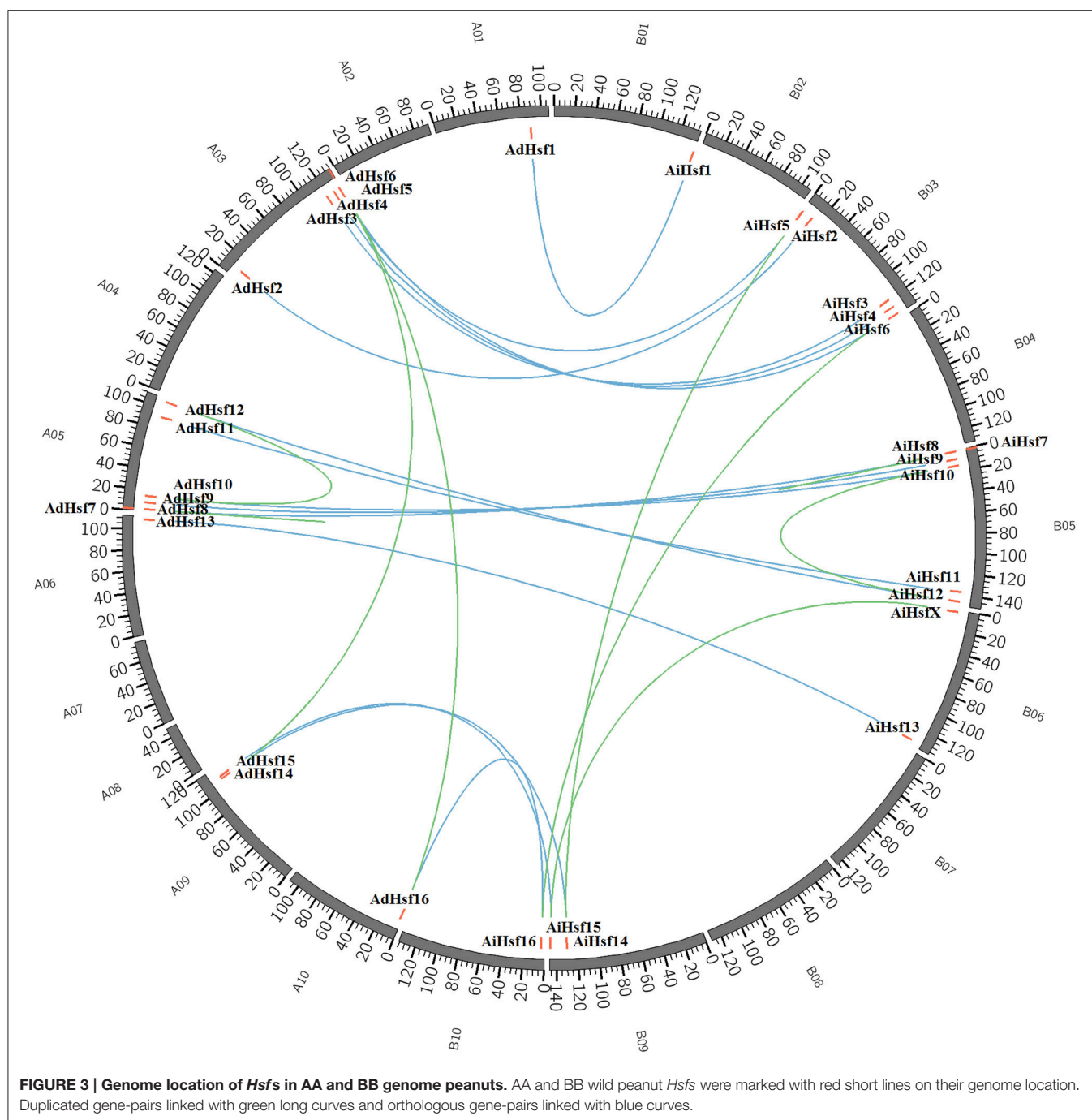
Expression of *Hsfs* in Various Tissues in Cultivated Peanut

We used *Hsfs* of wild peanut species as queries to identify *Hsfs* in cultivated peanut species from transcriptome and genomic sequences (unpublished data). Totally, 17 *Hsfs* were identified in cultivated peanut species and named as *AhHsf1*-*AhHsf16* and *AhHsfX*. The sequences of these genes were similar to their orthologous genes in wild peanut species (**Table S6**). To predict the possible function of these genes in cultivated peanut, the expression of these genes was investigated by qRT-PCR. Results showed that *AhHsf1*, 3, 7, 8, 11, 12, 14, 15, 16, and X were expressed predominantly in seeds, while the expression of *AhHsf9* and 10 was not detected in seeds. *AhHsf2*, 4, 5, 6, 9, and 10 were highly expressed in flower. The expression of *AhHsf1*, 7, 12, 15, and 16 was higher in flower than that in root, shoot or leaf. The expression of *AhHsf11*, 13, and X was higher in leaf than that in root, shoot or flower. The expression of *AhHsf4* and *AhHsf6* was higher in root than that in shoot or seed. The expression level of *AhHsf9* was higher in shoot than that in leaf or seed (**Figure 6**).

Hsf Expression in Response to Various Stresses in Cultivated Peanut

The expression of *AhHsf* was analyzed under high temperature, drought and low temperature by qRT-PCR. The expression levels of most *Hsfs* (*AhHsf1*, 3, 4, 5, 6, 7, 9, 10, 11, 13, 14, 15, and X) were up-regulated under high temperature. The expression of *AhHsf1*, 3, 9, 15, and X was up-regulated up to ~9-folds after 6 h treatment with 42°C. *AhHsf4*, 5, 6, 10, 11, and 13 could response rapidly to high temperature, and up-regulated after 1 h treatment. The expression of *AhHsf4*, 5, 6, 10, and 11 was continuously increased during 1–6 h of 42°C treatment. The expression of *AhHsf13* was decreased at 6 h after 42°C treatment (**Figure 7**).

The expression of most *AhHsfs* was up-regulated under drought stress. The expression levels of *AhHsf2*, 4, 5, 7, 12, 14, 15, and 16 were increased after 1 h of drought treatment. The expression of *AhHsf2*, 5, 12, 14, 15, and 16 was continuously increased during the first 6 h of drought treatment. The expression of *AhHsf1*, 3, 9, 10, and 11 was up-regulated after 6 h of drought stress (~15-folds). *AhHsfX* didn't respond much to drought stress (**Figure 8**).



The expression of most *AhHsfs* was up-regulated after 1 h of 4°C treatment, and then down-regulated at 6 h after treatment. The expression of *AhHsf12* was continuously up-regulated during 6 h of cold treatment. The expression of *AhHsf14* was decreased at 1 h and then increased at 6 h after 4°C treatment (Figure S3).

Previous study showed that *Hsfs* may be involved in disease resistance (Pick et al., 2012). In this study, we analyzed the expression of *AhHsfs* in peanut seeds after *A. flavus* infection. The expression of most *AhHsfs* was down-regulated in seed after

A. flavus inoculation, while the expression of *AhHsf2* and 14 was up-regulated (~1.5-fold; Figure 9).

DISCUSSIONS

Leguminous Contained Different *Hsf* Clusters

B5 cluster was not presented in Arabidopsis *Hsfs*, while B5 cluster was identified in most leguminous species, such as *C. cajan*, *L. japonicus* wild peanuts, and *G. max*. B5 *Hsf* cluster were not

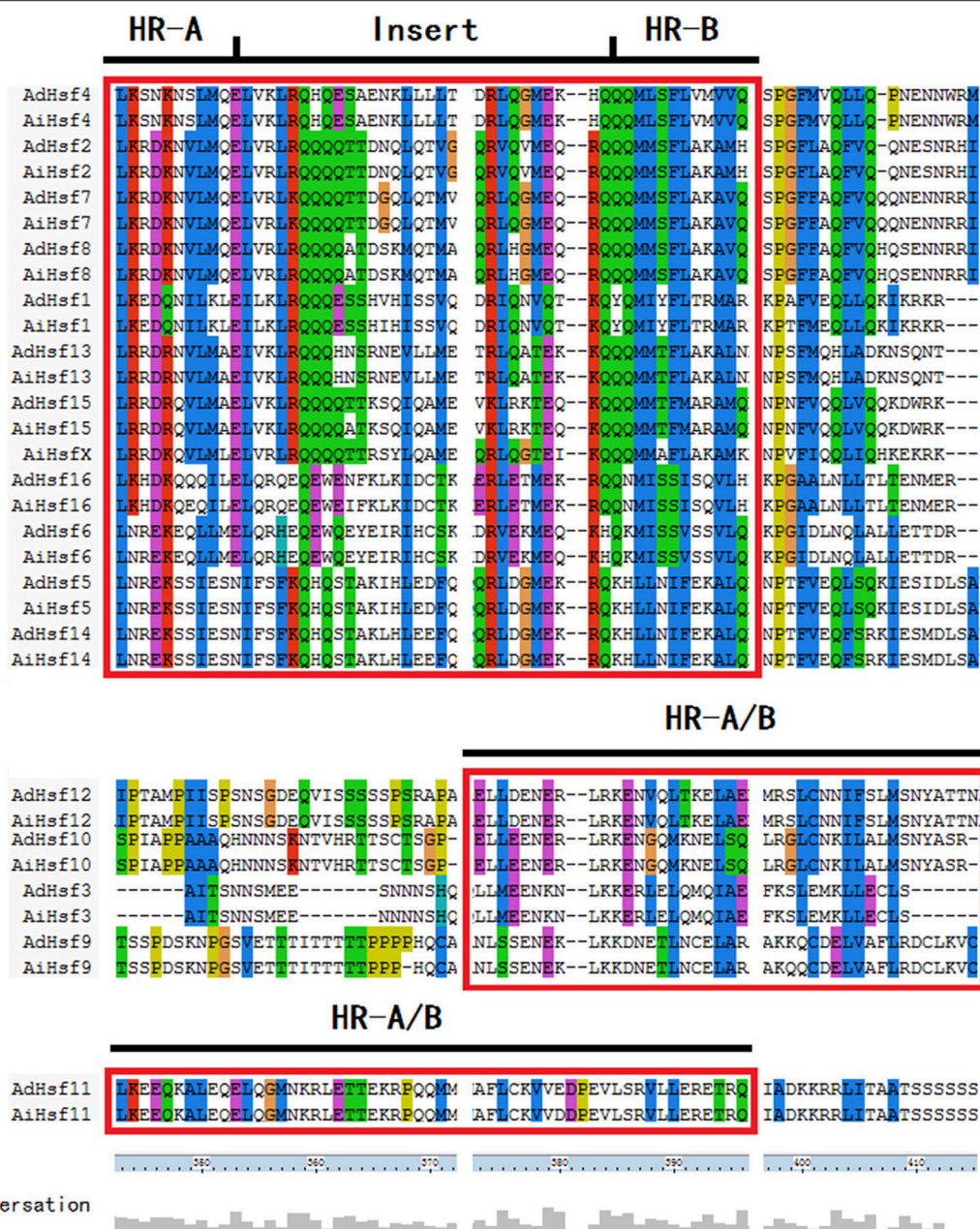


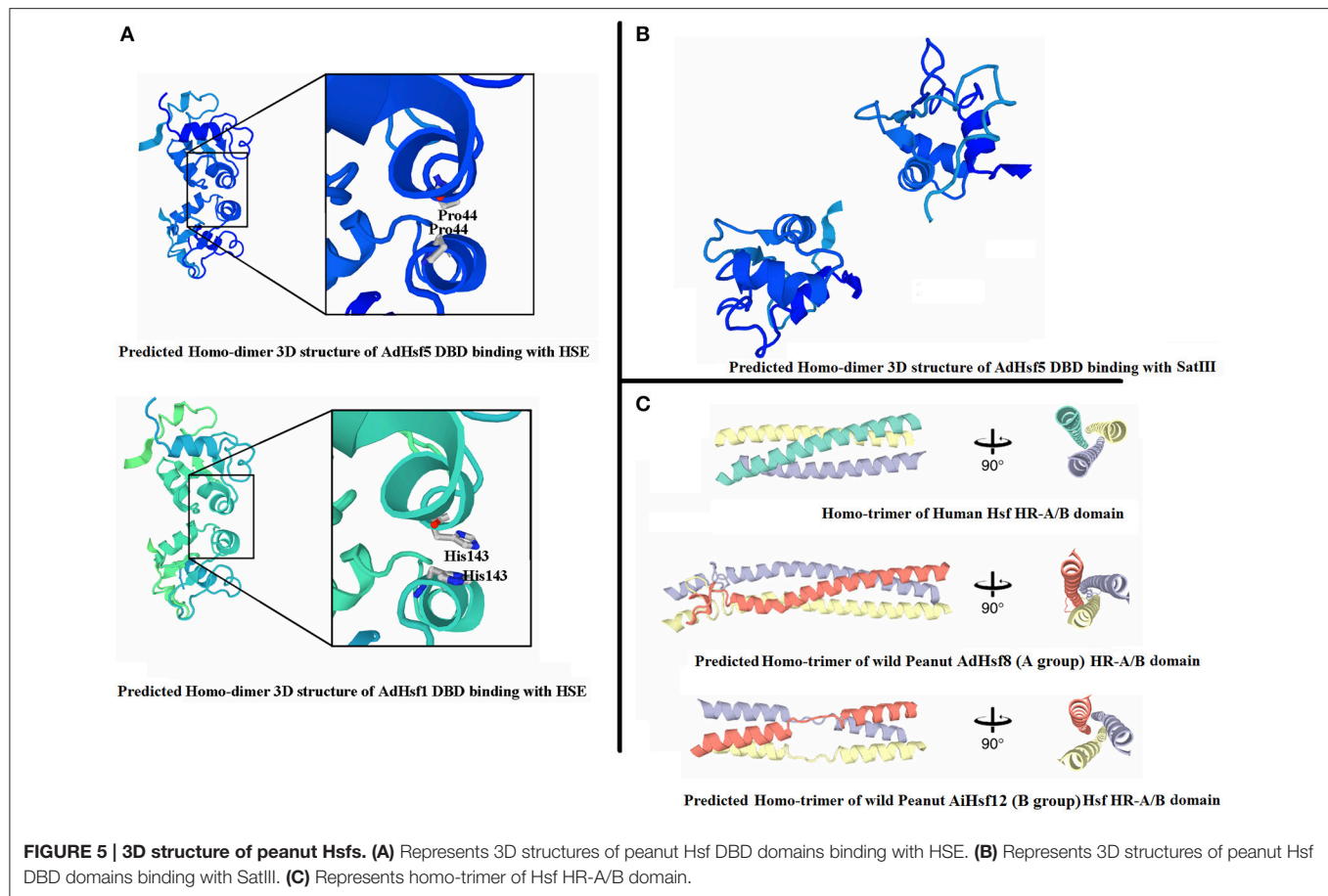
FIGURE 4 | HR-A/B domain in peanut Hsfs.

detected in *Medicago truncatula*. Phylogenetic tree showed that the leguminous plants contained different Hsf group members. Both in AA and BB wild peanut species, A3, A6a, A7, B3, and B4 cluster members were not found. Only soybean and *M. truncatula* contained the B3 members. A6a and A7 Hsf cluster was not found in leguminous. A3 cluster was not found in wild peanuts and *M. truncatula*. Group C Hsfs were not found in *L. japonicus* and *M. truncatula*. Soybean contained most clusters but not A6a and A7. The number of Hsfs from wild peanut species was relative small to compare with cotton, soybean and rosaceae (Li et al., 2014; Wang et al., 2014; Qiao et al., 2015). Phylogenetic tree showed

that *A. duranensis* is the closer relative of *A. ipaensis* compared with other Leguminous.

WGD may Not the Major Driving Force of Hsfs Large Scale Expansion in *Arachis*

Our results showed that Hsf gene duplication occurred in both AA and BB peanut genomes. The majority of Hsf duplication events were similar between AA and BB genomes. For example, the duplicated gene pair AdHsf7-AdHsf8 was located on chromosome 5. The distance between AdHsf7 and AdHsf8 was



about 1 kb. The duplicated gene pair *AiHsf7*-*AiHsf8* was located on chromosome 5, and the distance between these two genes was about 2 kb. However, *AiHsfX* was located on chromosome 6 of BB peanut and its duplicated gene *AiHsf15* was on chromosome 9 of BB peanut. It is possible that *AdHsf15* didn't undergo duplication or the orthologous of *AdHsfX* was lost during the evolution (Figure 3). We only found one tandem duplication gene pairs in *A. duranensis* and *A. ipaensis*, respectively. Both AA and BB genomes or their common ancestor were underwent the early papilionoid whole-genome duplication (WGD) about 58 million year ago ($K_s = 0.65$) (Bertioli et al., 2016). Intraspecies synteny analysis showed that *Hsf* duplication in wild peanut species was not originated from a large scale duplication event, because no intraspecies synteny blocks containing *Hsfs* was found. However, the recent WGD could be a driving force for the expansion of *Hsf* gene family in Chinese white pear and apple (Qiao et al., 2015). That may be the reason why peanut has less *Hsfs* than that in cotton, soybean and rosaceae (Li et al., 2014; Wang et al., 2014; Qiao et al., 2015).

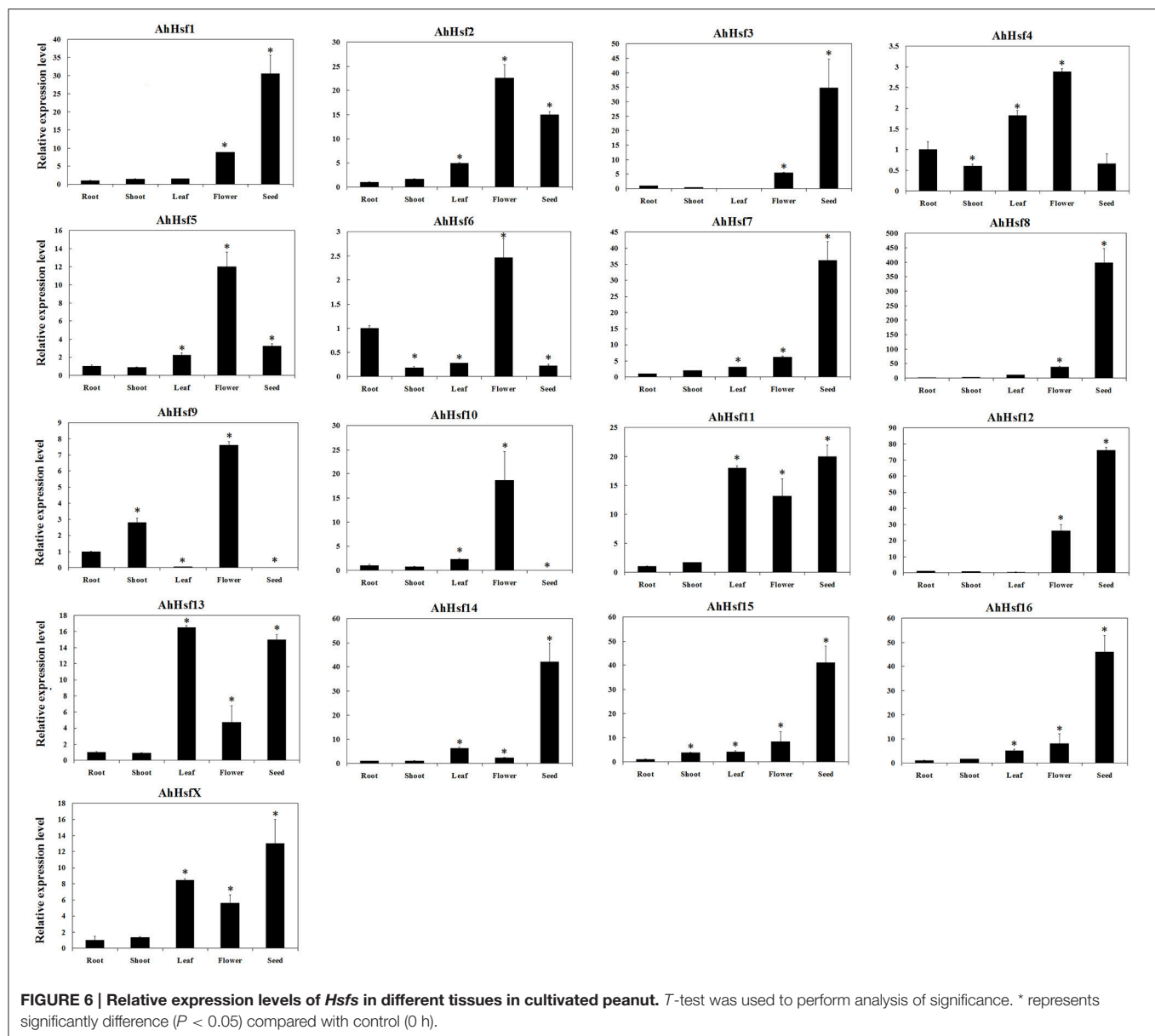
Hsfs Is Different in Group B from That in Group A

Group B *Hsfs* underwent positive selection (Table S4). Positive selection could contribute to adaptive evolution, functional

diversity, and neofunctionalization (Beisswanger and Stephan, 2008). Study on barley showed that many gene families involved in adaptation to environment were under positive selection. Positive selection may lead to the expansion of these gene families (Zeng et al., 2015). However, group A *Hsfs* underwent purifying selection. Purifying selection may generate genes with conserved functions or pseudogenization (Zhang, 2003). These results indicated that the function of *Arachis* group A *Hsfs* may be more conserved and the function of group B *Hsfs* may be more diverged. The sequences of *Hsf* group B HR-A/B were not conserved compare with group A HR-A/B which was in agreement with the differential selection they experienced (Figure 4). The 3D structure of peanut group B *Hsfs* was different from group A and C *Hsfs*. The 3D structure of group A and C HR-A/B was a continuous helix, while group B HR-A/B 3D structure contained helices which were linked by a linear part (Figure 5).

The Possible Roles of *Arachis Hsfs* in Abiotic and Biotic Stresses

Hsfs play a central role in protecting plants from high temperature or other stresses (Nishizawa-Yokoi et al., 2009; Scharf et al., 2012). Many *Hsfs* could regulate a set of heat-shock protein genes to enhance the thermo-tolerance in plants. Some *Hsfs* could be regulated by DREB genes as part of drought stress

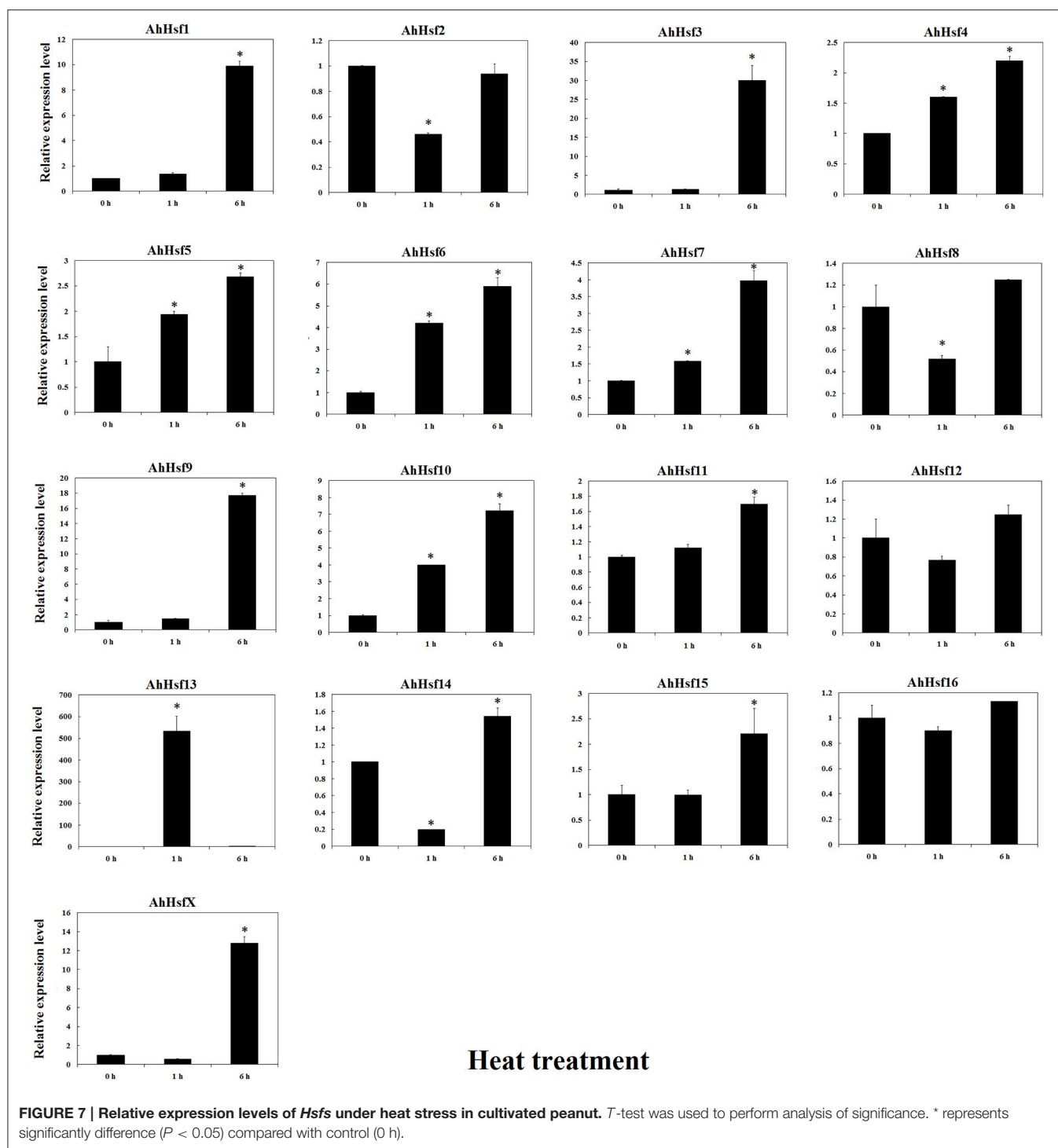


signaling pathway, and enhance drought tolerance (Scharf et al., 2012; Ma et al., 2015; Guo et al., 2016). Some Hsfs could regulate WRKY transcription factors which are involved in response to abiotic stresses, such as drought and cold (Ren et al., 2010; Zou et al., 2010; Jiang et al., 2012; Shen et al., 2015). Arabidopsis *HsfA9* could be activated by ABI3 to enhance seed desiccation tolerance and longevity (Verdier et al., 2013).

In our study, the majority of *Hsf* promoters contained HSE elements (Table S5) suggesting that peanut *Hsfs* could be regulated by other Hsfs. Many peanut *Hsf* promoters contained MYB binding sites which are involved in drought response (Table S5). It indicated that peanut *Hsfs* could be regulated by MYB transcription factors under drought stress. Many *Arachis Hsf* promoters contained ABRE and DRE elements which are involved in ABA-dependent or independent stress

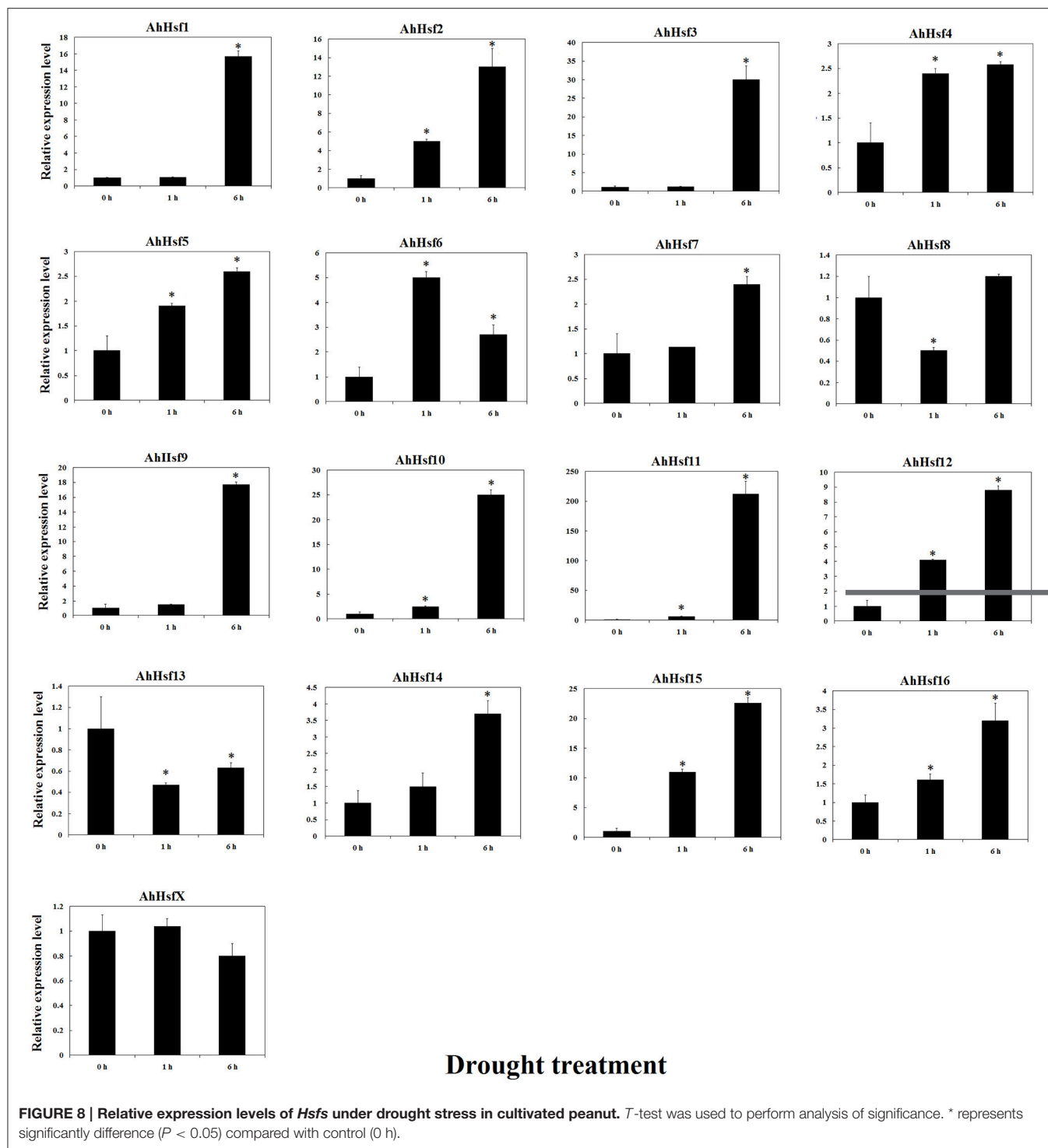
tolerance (Chen et al., 2012). Therefore, Hsfs could play important roles for gene regulation in response to different stresses in peanuts. Some *Arachis Hsf* promoters contained salicylic acid responsive, MeJA-responsive or fungal elicitor responsive elements, suggesting their roles in response to pathogen infection.

In cultivated peanut cultivars, the expression level of *AhHsf13* was approximately 500-folds as high as the control after 1 h of heat treatment, and then the expression was decreased after 6 h of treatment. Expression levels of *AhHsf1*, 3, 9, and *AhHsfX* were up-regulated by about 10-folds after 6 h of heat treatment to compare with the control. The expression of these *Hsfs* kept at a high level under continuous heat stress (Figure 7). Group A1a Hsfs were master regulators for acquired thermo tolerance in tomato and Arabidopsis (Scharf et al., 2012). However, we



found that the expression of AhHsf2 (group A1) did not respond to heat and cold, but to drought stress. In cultivated peanut, expression levels of AhHsf1, 2, 3, 9, 10, 11, 15 were about 10-folds as high to compare with the control after 6 h of drought stress (Figure 8). The expression of some Hsfs was altered after *Podosphaera aphanis* inoculation in woodland strawberry (Hu et al., 2015). *Aspergillus flavus* produces potent mycotoxin

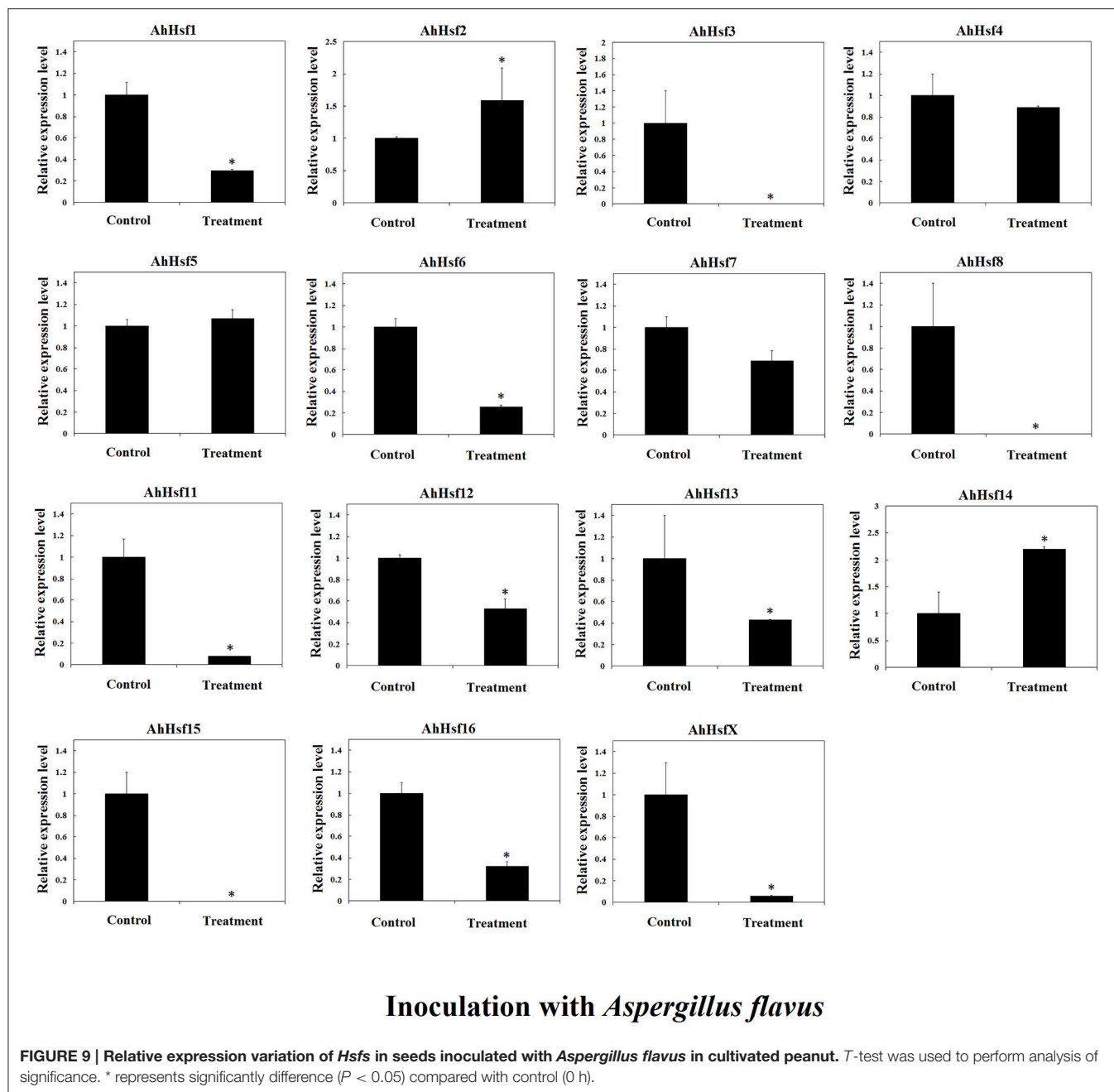
known as aflatoxin which is a key issue of food safety in peanut (Zhang et al., 2015). We detected whether peanut Hsf genes were involved in the response to *A. flavus* infection. The results showed that the expression of AhHsf2 and AhHsf14 were significantly up-regulated after *A. flavus* inoculation. The expression of some AhHsfs was down-regulated by *A. flavus* infection.



Hsf Gene Family Were Highly Expressed in Peanut Seed

Some *Hsfs* play key roles in plant seed development (Wang et al., 2014). In sunflower and *Arabidopsis*, *HsfA9* was expressed specifically in seeds and the expression of *Hsps* was changed during seed development (Almoguera et al.,

2002; Kotak et al., 2007b). In rice, *HsfA7* was expressed specifically in seed under normal condition (Chauhan et al., 2011). In peanut, expression levels of more than half of the *AhHsfs* were higher in seeds than that in other tissues. These expression patterns may suggest their roles in peanut seed development.



CONCLUSIONS

Genome-wide identification and comparison of peanut Hsfs with other plant species revealed that peanut contained a small number of Hsfs. Phylogenetic tree showed that B5 cluster Hsfs might present only in leguminous. Small scale segmental and tandem duplication but not WGD played important roles in Hsfs expansion in *Arachis*. The sequences of group B Hsf HR-A/B were not conserved compare with group A HR-A/B which was in agreement with the different selection pressure they experienced. We built the 3D structures of peanut Hsfs with the newly submitted templates and

found the difference between group A and B members. Peanut Hsfs may play important roles in abiotic and biotic stress tolerance based on their expression responses to these stresses.

AUTHOR CONTRIBUTIONS

XW designed the study, wrote the manuscript and finalized the figures and tables. PW and LH carried out the majority of experiments, data analysis, and wrote the method section of the manuscript. HS, CL, PL, AL, and HG performed experiments.

ACKNOWLEDGMENTS

This study was supported by the National grants (31500217; 31471526; 2013AA102602; 31500257), Shandong provincial grants (BS2013SW006; BS2014SW017; ZR2015CQ012; ZR2015YL061), Agricultural scientific, and technological innovation project (CXGC2016C08) and Young Talents Training Program of Shandong Academy of Agricultural Sciences.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00106/full#supplementary-material>

REFERENCES

- Akerfelt, M., Morimoto, R. I., and Sistonen, L. (2010). Heat shock factors: integrators of cell stress, development and lifespan. *Nat. Rev. Mol. Cell Biol.* 11, 545–555. doi: 10.1038/nrm2938
- Almoguera, C., Rojas, A., Diaz-Martin, J., Prieto-Dapena, P., Carranco, R., and Jordano, J. (2002). A seed-specific heat-shock transcription factor involved in developmental regulation during embryogenesis in sunflower. *J. Biol. Chem.* 277, 43866–43872. doi: 10.1074/jbc.M207330200
- Al-Whaibi, M. H. (2011). Plant heat-shock proteins: a mini review. *J. King Saud Univ. Sci.* 23, 139–150. doi: 10.1016/j.jksus.2010.06.022
- Baniwal, S. K., Bharti, K., Chan, K. Y., Fauth, M., Ganguli, A., Kotak, S., et al. (2004). Heat stress response in plants: a complex game with chaperones and more than twenty heat stress transcription factors. *J. Biosci.* 29, 471–487. doi: 10.1007/BF02712120
- Beisswanger, S., and Stephan, W. (2008). Evidence that strong positive selection drives neofunctionalization in the tandemly duplicated polyhomeotic genes in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 5447–5452. doi: 10.1073/pnas.0710892105
- Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–446. doi: 10.1038/ng.3517
- Brasileiro, A. C., Morgante, C. V., Araujo, A. C., Leal-Bertioli, S. C., Silva, A. K., Martins, A. C., et al. (2015). Transcriptome profiling of wild arachis from water-limited environments uncovers drought tolerance candidate genes. *Plant Mol. Biol. Rep.* 33, 1876–1892. doi: 10.1007/s11105-015-0882-x
- Chauhan, H., Khurana, N., Agarwal, P., and Khurana, P. (2011). Heat shock factors in rice (*Oryza sativa* L.): genome-wide expression analysis during reproductive development and abiotic stress. *Mol. Genet. Genomics* 286, 171–187. doi: 10.1007/s00438-011-0638-8
- Chen, L., Chen, Y., Jiang, J., Chen, S., Chen, F., Guan, Z., et al. (2012). The constitutive expression of *Chrysanthemum dichrum* ICE1 in *Chrysanthemum grandiflorum* improves the level of low temperature, salinity and drought tolerance. *Plant Cell Rep.* 31, 1747–1758. doi: 10.1007/s00299-012-1288-y
- Chung, E., Kim, K. M., and Lee, J. H. (2013). Genome-wide analysis and molecular characterization of heat shock transcription factor family in *Glycine max*. *J. Genet. Genomics* 40, 127–135. doi: 10.1016/j.jgg.2012.12.002
- Dalton, T. P., Solis, W. A., Nebert, D. W., and Carvan, M. J. (2000). Characterization of the MTF-1 transcription factor from zebrafish and trout cells. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 126, 325–335. doi: 10.1016/S0305-0491(00)00180-2
- Díaz-Martín, J., Almoguera, C., Prieto-Dapena, P., Espinosa, J. M., and Jordano, J. (2005). Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter. *Plant Physiol.* 139, 1483–1494. doi: 10.1104/pp.105.069963
- Döring, P., Treuter, E., Kistner, C., Lyck, R., Chen, A., and Nover, L. (2000). The role of AHA motifs in the activator function of tomato heat stress transcription factors HsfA1 and HsfA2. *Plant Cell* 12, 265–278. doi: 10.2307/3870927
- Freitas, F. O., Moretzsohn, M. C., and Valls, J. F. (2007). Genetic variability of Brazilian Indian landraces of *Arachis hypogaea* L. *Genet. Mol. Res.* 6, 675–684.
- Giorno, F., Guerriero, G., Baric, S., and Mariani, C. (2012). Heat shock transcriptional factors in *Malus domestica*: identification, classification and expression analysis. *BMC Genomics* 13:639. doi: 10.1186/1471-2164-13-639
- Grady, D. L., Ratliff, R. L., Robinson, D. L., McCanlies, E. C., Meyne, J., and Moyzis, R. K. (1992). Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. U.S.A.* 89, 1695–1699. doi: 10.1073/pnas.89.5.1695
- Guertin, M. J., Martins, A. L., Siepel, A., and Lis, J. T. (2012). Accurate prediction of inducible transcription factor binding intensities *in vivo*. *PLoS Genet.* 8:e1002610. doi: 10.1371/journal.pgen.1002610
- Guimarães, P. M., Brasileiro, A. C., Morgante, C. V., Martins, A. C., Pappas, G., Silva, O. B., et al. (2012). Global transcriptome analysis of two wild relatives of peanut under drought and fungi infection. *BMC Genomics* 13:387. doi: 10.1186/1471-2164-13-387
- Guo, M., Liu, J., Ma, X., Luo, D., Gong, Z., and Lu, M. (2016). The plant heat stress transcription factors (HSFs): structure, regulation and function in response to abiotic stresses. *Front. Plant Sci.* 7:114. doi: 10.3389/fpls.2016.00114
- Hartl, F. U., and Hayer-Hartl, M. (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* 295, 1852–1858. doi: 10.1126/science.1068408
- Hsu, A. L., Murphy, C. T., and Kenyon, C. (2003). Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* 300, 1142–1145. doi: 10.1126/science.1083701
- Hu, Y., Han, Y.-T., Wei, W., Li, Y.-J., Zhang, K., Gao, Y.-R., et al. (2015). Identification, isolation, and expression analysis of heat shock transcription factors in the diploid woodland strawberry *Fragaria vesca*. *Front. Plant Sci.* 6:736. doi: 10.3389/fpls.2015.00736
- Jaeger, A. M., Pemble, C. W., Sistonen, L., and Thiele, D. J. (2016). Structures of HSF2 reveal mechanisms for differential regulation of human heat-shock factors. *Nat. Struct. Mol. Biol.* 23, 147–154. doi: 10.1038/nsmb.3150
- Jiang, Y. J., Liang, G., and Yu, D. Q. (2012). Activated expression of WRKY57 confers drought tolerance in *Arabidopsis*. *Mol. Plant* 5, 1375–1388. doi: 10.1093/mp/sss080
- Kochert, G., Stalker, H. T., Gimenes, M., Galgaro, L., Romero-Lopes, C., and Moore, K. (1996). RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *Am. J. Bot.* 83, 1282–1291.
- Kotak, S., Larkindale, J., Lee, U., von Koskull-Döring, P., Vierling, E., and Scharf, K.-D. (2007a). Complexity of the heat stress response in plants. *Curr. Opin. Plant Biol.* 10, 310–316. doi: 10.1016/j.pbi.2007.04.011

Figure S1 | Structure of peanut *Hsfs*.

Figure S2 | DBD domain in peanut *Hsfs*.

Figure S3 | Relative expression levels of *Hsfs* under cold stress in cultivated peanut. *T*-test was used to perform analysis of significance. “*” represents significantly difference ($P < 0.05$) compared with control (0 h).

Table S1 | *Hsfs* identified in other species.

Table S2 | Primers of cultivated peanut *Hsfs* for qRT-PCR.

Table S3 | Identity of wild peanut *Hsf* orthologous.

Table S4 | Likelihood values and parameter estimate for wild peanut *Hsfs*.

Table S5 | Cis-acting regulatory elements of wild peanut *Hsf* promoters.

Table S6 | Identity of wild and cultivated peanut *Hsfs*.

- Kotak, S., Port, M., Ganguli, A., Bicker, F., and von Koskull-Döring, P. (2004). Characterization of C-terminal domains of Arabidopsis heat stress transcription factors (Hsfs) and identification of a new signature combination of plant class A Hsfs with AHA and NES motifs essential for activator function and intracellular localization. *Plant J.* 39, 98–112. doi: 10.1111/j.1365-313X.2004.02111.x
- Kotak, S., Vierling, E., Bäumlein, H., and von Koskull-Döring, P. (2007b). A novel transcriptional cascade regulating expression of heat stress proteins during seed development of Arabidopsis. *Plant Cell* 19, 182–195. doi: 10.1105/tpc.106.048165
- Kreps, J. A., Wu, Y., Chang, H. S., Zhu, T., Wang, X., and Harper, J. F. (2002). Transcriptome changes for Arabidopsis in response to salt, osmotic, and cold stress. *Plant Physiol.* 130, 2129–2141. doi: 10.1104/pp.008532
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Li, P. S., Yu, T. F., He, G. H., Chen, M., Zhou, Y. B., Chai, S. C., et al. (2014). Genome-wide analysis of the Hsf family in soybean and functional identification of GmHsf-34 involvement in drought and heat stresses. *BMC Genomics* 15:1009. doi: 10.1186/1471-2164-15-1009
- Lin, Y., Cheng, Y., Jin, J., Jin, X., Jiang, H., Yan, H., et al. (2014). Genome duplication and gene loss affect the evolution of heat shock transcription factor genes in legumes. *PLoS ONE* 9:e102825. doi: 10.1371/journal.pone.0102825
- Ma, H., Wang, C., Yang, B., Cheng, H., Wang, Z., Mijiti, A., et al. (2015). CarHSFB2, a Class B heat shock transcription factor, is involved in different developmental processes and various stress responses in chickpea (*Cicer arietinum* L.). *Plant Mol. Biol. Rep.* 34, 1–14. doi: 10.1007/s11105-015-0892-8
- Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M., et al. (2005). Modeling gene and genome duplications in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 5454–5459. doi: 10.1073/pnas.0501102102
- Mendillo, M. L., Santagata, S., Koeva, M., Bell, G., Rong, H., Tamimi, R. M., et al. (2012). HSF1 drives a transcriptional program distinct from heat shock to support highly malignant human cancers. *Cell* 150, 549–562. doi: 10.1016/j.cell.2012.06.031
- Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C. M., Valls, J. F. M., and Bertioli, D. J. (2013). A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot.* 111, 113–126. doi: 10.1093/aob/mcs237
- Neudegger, T., Verghese, J., Hayer-Hartl, M., Hartl, F. U., and Bracher (2016). A. Structure of human heat-shock transcription factor 1 in complex with DNA. *Nat. Struct. Mol. Biol.* 23, 140–146. doi: 10.1038/nsmb.3149
- Nishizawa-Yokoi, A., Yoshida, E., Yabuta, Y., and Shigeoka, S. (2009). Analysis of the regulation of target genes by an Arabidopsis heat shock transcription factor, HsfA2. *Biosci. Biotechnol. Biochem.* 73, 890–895. doi: 10.1271/bbb.80809
- Nover, L., Bharti, K., Döring, P., Mishra, S. K., Ganguli, A., and Scharf, K.-D. (2001). Arabidopsis and the heat stress transcription factor world: how many heat stress transcription factors do we need? *Cell Stress Chaperones* 6, 177–189.
- Pelham, H. R. B. (1982). A regulatory upstream promoter element in the *Drosophila* hsp70 heat-shock gene. *Cell* 30, 517–528. doi: 10.1016/0092-8674(82)90249-5
- Peteranderl, R., Rabenstein, M., Shin, Y. K., Liu, C. W., Wemmer, D. E., King, D. S., et al. (1999). Biochemical and biophysical characterization of the trimerization domain from the heat shock transcription factor. *Biochemistry* 38, 3559–3569. doi: 10.1021/bi981774j
- Pick, T., Jaskiewicz, M., Peterhansel, C., and Conrath, U. (2012). Heat shock factor HsfB1 primes gene transcription and systemic acquired resistance in Arabidopsis. *Plant Physiol.* 159, 52–55. doi: 10.1104/pp.111.191841
- Qiao, X., Li, M., Li, L., Yin, H., Wu, J., and Zhang, S. (2015). Genome-wide identification and comparative analysis of the heat shock transcription factor family in Chinese white pear (*Pyrus bretschneideri*) and five other Rosaceae species. *BMC Plant Biol.* 15:12. doi: 10.1186/s12870-014-0401-5
- Ramu, V. S., Swetha, T. N., Sheela, S. H., Babitha, C. K., Rohini, S., Reddy, M. K., et al. (2015). Simultaneous expression of regulatory genes associated with specific drought-adaptive traits improves drought adaptation in peanut. *Plant Biotechnol. J.* 14, 1008–1020. doi: 10.1111/pbi.12461
- Ren, X. Z., Chen, Z. Z., Liu, Y., Zhang, H. R., Zhang, M., Liu, Q. A., et al. (2010). ABO3, a WRKY transcription factor, mediates plant responses to abscisic acid and drought tolerance in Arabidopsis. *Plant J.* 63, 417–429. doi: 10.1111/j.1365-313X.2010.04248.x
- Ritossa, F. (1962). A new puffing pattern induced by temperature shock and DNP in *Drosophila*. *Experientia* 18, 571–573. doi: 10.1007/BF02172188
- Rizzon, C., Ponger, L., and Gaut, B. S. (2006). Striking similarities in the genomic distribution of tandemly arrayed genes in Arabidopsis and rice. *PLoS Comput. Biol.* 2:e115. doi: 10.1371/journal.pcbi.0020115
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15, 227–239. doi: 10.1093/dnares/dsn008
- Scharf, K. D., Berberich, T., Ebersberger, I., and Nover, L. (2012). The plant heat stress transcription factor (Hsf) family: structure, function and evolution. *Biochim. Biophys. Acta* 1819, 104–119. doi: 10.1016/j.bbagr.2011.10.002
- Schmidt, R., Schippers, J. H., Welker, A., Mieulet, D., Guiderdoni, E., and Mueller-Roeber, B. (2012). Transcription factor OsHsfC1b regulates salt tolerance and development in *Oryza sativa* ssp. japonica. *AoB Plants* 2012:pls011. doi: 10.1093/aobpla/pls011
- Schwechheimer, C., and Bevan, M. (1998). The regulation of transcription factor activity in plants. *Trends Plant Sci.* 3, 378–383. doi: 10.1016/S1360-1385(98)01302-8
- Shen, Z., Yao, J., Sun, J., Chang, L., Wang, S., Ding, M., et al. (2015). Populus euphratica HSF binds the promoter of WRKY1 to enhance salt tolerance. *Plant Sci.* 235, 89–100. doi: 10.1016/j.plantsci.2015.03.006
- Shinozaki, K., and Yamaguchi-Shinozaki, K. (2007). Gene networks involved in drought stress response and tolerance. *J. Exp. Bot.* 58, 221–227. doi: 10.1093/jxb/erl164
- Swindell, W. R., Huebner, M., and Weber, A. P. (2007). Transcriptional profiling of Arabidopsis heat shock proteins and transcription factors reveals extensive overlap between heat and non-heat stress response pathways. *BMC Genomics* 8:125. doi: 10.1186/1471-2164-8-125
- Tanabe, M., Nakai, A., Kawaoe, Y., and Nagata, K. (1997). Different thresholds in the responses of two heat shock transcription factors, HSF1 and HSF3. *J. Biol. Chem.* 272, 15389–15395. doi: 10.1074/jbc.272.24.15389
- Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K., et al. (2013). A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiol.* 163, 757–774. doi: 10.1104/pp.113.222380
- Vihervara, A., Sergelius, C., Vasara, J., Blom, M. A. H., Elsing, A. N., Roos-Mattjus, P., et al. (2013). Transcriptional response to stress in the dynamic chromatin environment of cycling and mitotic cells. *Proc. Natl. Acad. Sci. U.S.A.* 110, E3388–E3397. doi: 10.1073/pnas.1305275110
- Von Koskull-Döring, P., Scharf, K.-D., and Nover, L. (2007). The diversity of plant heat stress transcription factors. *Trends Plant Sci.* 12, 452–457. doi: 10.1016/j.tplants.2007.08.014
- Wang, F., Dong, Q., Jiang, H., Zhu, S., Chen, B., and Xiang, Y. (2012). Genome-wide analysis of the heat shock transcription factors in *Populus trichocarpa* and *Medicago truncatula*. *Mol. Biol. Rep.* 39, 1877–1886. doi: 10.1007/s11033-011-0933-9
- Wang, J., Sun, N., Deng, T., Zhang, L., and Zuo, K. (2014). Genome-wide cloning, identification, classification and functional analysis of cotton heat shock transcription factors in cotton (*Gossypium hirsutum*). *BMC Genomics* 15:961. doi: 10.1186/1471-2164-15-961
- Wang, P., Gao, C., Bian, X., Zhao, S., Zhao, C., Xia, H., et al. (2016). Genome-wide identification and comparative analysis of cytosine-5 DNA methyltransferase and demethylase families in wild and cultivated peanut. *Front. Plant Sci.* 7:7. doi: 10.3389/fpls.2016.00007
- Wang, W., Vinocur, B., Shoseyov, O., and Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci.* 9, 244–252. doi: 10.1016/j.tplants.2004.03.006
- Xue, G. P., Sadat, S., Drenth, J., and McIntyre, C. L. (2014). The heat shock factor family from *Triticum aestivum* in response to heat and other major abiotic stresses and their role in regulation of heat shock protein genes. *J. Exp. Bot.* 65, 539–557. doi: 10.1093/jxb/ert399
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Yang, Z., Wang, Y., Gao, Y., Zhou, Y., Zhang, E., Hu, Y., et al. (2014). Adaptive evolution and divergent expression of heat stress transcription factors in grasses. *BMC Evol. Biol.* 14:147. doi: 10.1186/1471-2148-14-147
- Yuan, S. X., Xu, B., Zhang, J., Xie, Z. N., Cheng, Q., Yang, Z., et al. (2015). Comprehensive analysis of CCCH-type zinc finger family genes facilitates functional gene discovery and reflects recent allopolyploidization event in tetraploid switchgrass. *BMC Genomics* 16:129. doi: 10.1186/s12864-015-1328-4
- Zeng, X., Long, H., Wang, Z., Zhao, S., Tang, Y., Huang, Z., et al. (2015). The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. *Proc. Natl. Acad. Sci. U.S.A.* 112, 1095–1100. doi: 10.1073/pnas.1423628112
- Zhang, H., Scharfenstein, L., Zhang, D., Chang, P., Montalbano, B. G., Guo, B., et al. (2015). Peanut resistance gene expression in response to *Aspergillus flavus* infection during seed germination. *J. Phytopathol.* 163, 212–221. doi: 10.1111/jph.12311
- Zhang, J. Z. (2003). Evolution by gene duplication: an update. *Trends Ecol.* 18, 292–298. doi: 10.1016/s0169-5347(03)00033-8
- Zhang, X., Feng, Y., Cheng, H., Tian, D., and Yang, S. (2011). Relative evolutionary rates of NBS-encoding genes revealed by soybean segmental duplication. *Mol. Genet. Genomics* 285, 79–90. doi: 10.1007/s00438-010-0587-7
- Zou, C. S., Jiang, W. B., and Yu, D. Q. (2010). Male gametophyte-specific WRKY34 transcription factor mediates cold sensitivity of mature pollen in *Arabidopsis*. *J. Exp. Bot.* 61, 3901–3914. doi: 10.1093/jxb/erq204

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Wang, Song, Li, Li, Li, Guan, Hou and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Overexpression of *Nictaba*-Like Lectin Genes from *Glycine max* Confers Tolerance toward *Pseudomonas syringae* Infection, Aphid Infestation and Salt Stress in Transgenic *Arabidopsis* Plants

Sofie Van Holle¹, Guy Smagghe² and Els J. M. Van Damme^{1*}

¹ Laboratory of Biochemistry and Glycobiology, Department of Molecular Biotechnology, Ghent University, Ghent, Belgium,

² Laboratory of Agrozoology, Department of Crop Protection, Ghent University, Ghent, Belgium

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Kathryn Kamo,
United States Department of
Agriculture, USA
Milena Roux,
University of Copenhagen, Denmark

*Correspondence:

Els J. M. Van Damme
elsjm.vandamme@ugent.be

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 07 September 2016

Accepted: 07 October 2016

Published: 25 October 2016

Citation:

Van Holle S, Smagghe G and Van
Damme EJM (2016) Overexpression
of *Nictaba*-Like Lectin Genes from
Glycine max Confers Tolerance
toward *Pseudomonas syringae*
Infection, Aphid Infestation and Salt
Stress in Transgenic *Arabidopsis*
Plants. *Front. Plant Sci.* 7:1590.
doi: 10.3389/fpls.2016.01590

Plants have evolved a sophisticated immune system that allows them to recognize invading pathogens by specialized receptors. Carbohydrate-binding proteins or lectins are part of this immune system and especially the lectins that reside in the nucleocytoplasmic compartment are known to be implicated in biotic and abiotic stress responses. The class of *Nictaba*-like lectins (NLL) groups all proteins with homology to the tobacco (*Nicotiana tabacum*) lectin, known as a stress-inducible lectin. Here we focus on two *Nictaba* homologs from soybean (*Glycine max*), referred to as *GmNLL1* and *GmNLL2*. Confocal laser scanning microscopy of fusion constructs with the green fluorescent protein either transiently expressed in *Nicotiana benthamiana* leaves or stably transformed in tobacco BY-2 suspension cells revealed a nucleocytoplasmic localization for the *GmNLLs* under study. RT-qPCR analysis of the transcript levels for the *Nictaba*-like lectins in soybean demonstrated that the genes are expressed in several tissues throughout the development of the plant. Furthermore, it was shown that salt treatment, *Phytophthora sojae* infection and *Aphis glycines* infestation trigger the expression of particular *NLL* genes. Stress experiments with *Arabidopsis* lines overexpressing the *NLLs* from soybean yielded an enhanced tolerance of the plant toward bacterial infection (*Pseudomonas syringae*), insect infestation (*Myzus persicae*) and salinity. Our data showed a better performance of the transgenic lines compared to wild type plants, indicating that the *NLLs* from soybean are implicated in the stress response. These data can help to further elucidate the physiological importance of the *Nictaba*-like lectins from soybean, which can ultimately lead to the design of crop plants with a better tolerance to changing environmental conditions.

Keywords: lectin, *Nictaba*, soybean, *Phytophthora sojae*, *Pseudomonas syringae*, *Myzus persicae*, *Aphis glycines*, salt stress

INTRODUCTION

To successfully survive in their natural habitat, plants are capable of experiencing stress when they are confronted with adverse environmental conditions including drought, insect infestation or pathogen infection. Because plants cannot flee from these unfavorable conditions, they have developed a sophisticated protection system which enables them to recognize disadvantageous situations, alter hormone crosstalk and successfully cope with these adverse growth conditions (Jones and Dangl, 2006). The plant's innate immune system can recognize invading pathogens by a range of specialized cell-surface and intracellular receptors. It was shown that lectins are part of the plant's immune system since they can act as immune receptors and/or defense proteins (Peumans and Van Damme, 1995; Lannoo and Van Damme, 2014).

The class of plant carbohydrate-binding proteins or lectins is widespread within the plant kingdom and these proteins exhibit specificities toward endogenous as well as exogenous glycan structures (Van Damme et al., 2008). During the last decade, compelling evidence has been offered demonstrating that next to the classical lectins that reside mostly in the vacuole, there is a group of inducible cytoplasmic/nuclear lectins. The latter group of lectins is not easily detectable in plants under normal environmental conditions, but their expression level is increased after application of certain stressors (Van Damme et al., 2004; Lannoo and Van Damme, 2010). At present, at least six carbohydrate recognition domains have been identified within the group of nucleocytoplasmic lectins (Lannoo and Van Damme, 2010). Several of these nucleocytoplasmic lectins have been studied in detail and play roles in plant stress signaling (Al Atalah et al., 2014; Van Hove et al., 2015). One of these domains was first discovered in the *Nicotiana tabacum* (tobacco) agglutinin, abbreviated as Nictaba (Chen et al., 2002). In recent years, Nictaba was also shown to be implicated in the plant stress response (Chen et al., 2002; Lannoo et al., 2007; Vandenborre et al., 2009a, 2010; Delporte et al., 2011). This GlcNAc-binding lectin is believed to trigger gene expression in response to stress by interaction with the core histones H2A, H2B and H4 through their O-GlcNAc modification (Schouppe et al., 2011; Delporte et al., 2014).

An extensive survey of genome databases revealed that Nictaba-like lectins (NLLs) are widespread in plants (Delporte et al., 2015). Thus, far, functional characterization has been focused on the tobacco lectin and one F-box Nictaba homolog from *Arabidopsis* (Stefanowicz et al., 2012; Delporte et al., 2015). Lectin expression in tobacco is enhanced after caterpillar attack, suggesting a role for Nictaba in plant defense. Furthermore, experiments using transgenic tobacco plants overexpressing the lectin gene or plants with reduced expression indicated that Nictaba exerts insecticidal activity toward Lepidopteran

pest insects (Vandenborre et al., 2010). The *Arabidopsis* F-box-Nictaba homolog is upregulated after treatment with salicylic acid and upon *Pseudomonas syringae* infection and overexpression of the gene in *Arabidopsis* plants confers increased tolerance to the pathogen (Stefanowicz et al., 2016). In order to refine our understanding of this specific group of nucleocytoplasmic lectins, we focus here on some Nictaba-like lectins from soybean. Soybean presents an exciting opportunity to investigate the stress inducibility of these proteins in an important crop species. Several *GmNictaba*-related genes have recently been identified in the soybean genome. Of the 31 identified *GmNLL* genes, 25 encode chimerolectins, consisting of one Nictaba lectin domain combined with an N-terminal F-box protein domain. The remaining six genes encode Nictaba orthologs containing one or two Nictaba domains as building blocks (Van Holle and Van Damme, 2015).

In this study, two *GmNLL* genes, referred to as *GmNLL1* and *GmNLL2*, located on different chromosomes have been selected for analysis. Their localization in the cell was investigated, together with their temporal and spatial expression in wild type soybean plants subjected to a variety of abiotic and biotic stresses. In addition, *Arabidopsis* overexpression lines were generated and analyzed for tolerance toward pathogen infection and aphid infestation. These data allowed us to investigate if overexpression of the *GmNictaba*-related genes leads to an enhanced tolerance of the plant toward stress.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Wild type seeds of *Arabidopsis thaliana* ecotype Colombia were purchased from Lehle Seeds (Texas, USA). For *in vitro* cultures, seeds were surface sterilized by submergence in 70% ethanol for 2 min, followed by 10 min in 5% NaOCl. Finally, the seeds were rinsed four to five times with sterilized water. *In vitro* cultures were maintained in a plant growth room at 21°C and a 16/8 h light/dark photoperiod. *Arabidopsis* plants were sown into Jiffy-7® (artificial soil) and grown in a Conviron (Berlin, Germany) plant growth cabinet under 12/12 h light/dark conditions at 21°C after stratification at 4°C for 3 days. Seeds for the insect assays were sown in round plastic pots (diameter: 11 cm) containing soil. After stratification pots were moved to a plant growth incubator (MLR-352 incubator, Sanyo/Panasonic, Osaka, Japan, 21°C, 12 h photoperiod, 75% relative humidity).

Glycine max cv Williams seeds were obtained from the USDA Soybean Germplasm Collection in Urbana (IL, USA). *Glycine max* cv Opaline seeds were obtained from the Institute for Agricultural and Fisheries Research (Merelbeke, Belgium). Seeds were grown in pots containing a mixture (50/50) of commercial soil and expanded clay granules (Agrex) in a growth chamber at 26°C with a 16/8 h light/dark photoperiod.

Nicotiana benthamiana seeds were kindly supplied by dr. Verne A. Sisson (Oxford Tobacco Research Station, Oxford, NC, USA). *N. benthamiana* plants were sown in pots containing commercial soil and grown in a growth chamber at 26°C with a 16/8 h light/dark photoperiod. The *N. tabacum* cv Bright Yellow-2 (BY-2) cell suspension culture was obtained from

Abbreviations: ABA, abscisic acid; BY-2, Bright Yellow-2; EGFP, enhanced green fluorescent protein; MeJA, methyl jasmonate; Murashige and Skoog, MS; NLL, Nictaba-like lectin; SA, salicylic acid; SBA, soybean agglutinin; SVL, soybean vegetative lectin.

the department of Plant Systems Biology (Flanders Institute for Biotechnology, Zwijnaarde, Belgium) and maintained as described by Delporte et al. (2014).

Pathogens

Phytophthora sojae was obtained from the CBS-KNAW Fungal Biodiversity Centre (Utrecht, The Netherlands) and was routinely cultured on 10% clarified and buffered V8-juice agar plates at 21°C in the dark. *Phytophthora brassicae* was grown under the same conditions and was kindly provided by Prof. Monica Höfte (Dept. of Crop Protection, Ghent University). *Pseudomonas syringae* pv. *tomato* strain DC3000 was also provided by Prof. Monica Höfte (Dept. of Crop Protection, Ghent University) and grown on King's B agar medium supplemented with 50 µg/ml rifampicin.

Cloning of the *Nictaba*-Like Sequences from Soybean

Trifoliate leaves from 18-day-old soybean (*Glycine max* cv Williams) plants were collected for RNA extraction. Total RNA was extracted using TRI Reagent® according to the manufacturer's instructions (Sigma-Aldrich). Residual genomic DNA was removed by a DNase I treatment (Life Technologies, Carlsbad, CA, USA) and RNA was quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Reverse transcriptase reactions were performed with 1 µg of total RNA using moloney murine leukemia virus reverse transcriptase (M-MLV RT) and oligo(dT)25 primers (Life Technologies). The full length cDNA sequences corresponding to *NLL1* (*Glyma.06G221100*) and *NLL2* (*Glyma.20G020900*) were obtained by RT-PCR reactions with gene specific primers (Supplementary Table 1). Finally, the PCR products were ligated in the pJET2.1 vector with the CloneJET PCR Cloning kit according to the manufacturer's instructions (Life Technologies) and constructs were sequenced (LGC Genomics, Berlin, Germany) to confirm the cDNA sequence of the *GmNLL* genes.

Construction of Expression Vectors

Vectors for expression of each of the *GmNLL* sequences either N- or C-terminally linked to EGFP (enhanced green fluorescent protein) under control of the CaMV 35S promoter were constructed using Life Technologies' Gateway® Cloning Technology. First, the cDNA clones were used as template in two consecutive PCRs and amplified with primers to attach *attB* sites to the PCR product. In the first PCR, the coding sequence of the *GmNLLs* was amplified using Platinum® Pfx DNA Polymerase (Life Technologies) and primers with stop codon (evd1022/evd1032 (*NLL1*) and evd1024/evd1033 (*NLL2*)) or without stop codon (evd1022/evd1023 (*NLL1*) and evd1024/evd1025 (*NLL2*)) (Supplementary Table 2) using the following cycling parameters: 2 min at 94°C, 25 cycles (15 s at 94°C, 30 s at 48°C, 1.5 min at 68°C), 5 min at 68°C. In the second PCR primers evd2/evd4 were used to complete the *attB* sites using following cycling parameters: 2 min at 94°C, 5 cycles (15 s at 94°C, 30 s at 48°C, 1.5 min at 68°C), 25 cycles (15 s at 94°C, 30 s at 55°C, 1.5 min at 68°C), 5 min at 68°C. The PCR products were used as substrates in a BP recombination reaction

with the pDONR221 donor vector. Subsequently, the entry clones were recombined with destination vectors pK7WGF2,0 and pK7FWG2,0 to create the desired expression clones to create N- or C-terminal EGFP fusions to the *NLL* gene sequences, respectively (Karimi et al., 2002). Using a similar approach, coding sequences of *GmNLL1* and *GmNLL2* were introduced into the binary vector pK7WG2,0 (Karimi et al., 2002) to generate expression vectors for transformation of *Arabidopsis* plants.

The binary vectors carrying the different constructs were introduced into *Agrobacterium tumefaciens* C58C1 Rif (pGV4000) using the freeze/thaw transformation method. Briefly, 1 µg of the expression clones was added to competent *A. tumefaciens* cells followed by an incubation of 30 min on ice. Next, the cells were frozen in liquid nitrogen, thawed at 37°C for 5 min, and after addition of 1 ml of preheated LB medium, the cells were incubated for 2 h at 26°C. Transformed cells were selected on LB agar plates containing 50 µg/ml spectinomycin and screened by colony PCR.

Transformation of *N. benthamiana* Plants and *N. tabacum* cv BY-2 Cells

Transient expression of the EGFP fusion proteins was conducted as described by Sparkes et al. (2006). The abaxial epidermis of young leaves of 4- to 6-week-old *N. benthamiana* plants was infiltrated with the *Agrobacterium* suspension harboring the different constructs. Two days post-infiltration, the infiltrated leaf areas were cut and analyzed microscopically. The tobacco BY-2 cell suspension culture was stably transformed with the EGFP fusion constructs under the control of the 35S promoter as described by Delporte et al. (2014).

Generation of *Arabidopsis* Transgenic Lines

Arabidopsis 35S::*GmNLL1* and 35S::*GmNLL2* overexpression lines were generated using the floral dip method (Clough and Bent, 1998). Transformed seeds were selected using the adapted protocol proposed by Harrison et al. (2006). Integration of the T-DNA was detected by RT-PCR on cDNA with gene specific primers (Supplementary Table 3) using the following PCR program: 5 min at 95°C, 40 cycles of 45 s at 95°C, 45 s at 60°C, and 30 s at 72°C and a final 5 min at 72°C. Relative expression levels of the *GmNLL* genes were analyzed in 4-week-old plants by RT-qPCR. At least three independent homozygous single insertion lines of 35S::*GmNLL1* and 35S::*GmNLL2* were selected and used in all experiments, together with the corresponding wild type plant.

Hormone Treatment and Abiotic Stress Application of Wild Type Soybean Plants

For hormone and salt stress treatments, 14-day-old soybean (*Glycine max* cv Williams) plants (V1 growth stage) were carefully removed from the soil and transferred to liquid Murashige and Skoog (MS) medium containing different hormones (100 µM abscisic acid (ABA), 50 µM methyl jasmonate (MeJA) or 300 mM salicylic acid (SA)) or 150 mM NaCl. For control treatments, equal volumes of the dissolvent

(ethanol or water) of the hormone or salt solution were added to the medium. Treated root and shoot tissues were sampled at the following time points: 3, 6, 10, 24, and/or 32 h. Likewise, the corresponding mock controls were sampled at each time point. Plant material of four individual plants was pooled for each sample and immediately frozen in liquid nitrogen and stored at -80°C until use. In total, three biological replicates were performed.

Infection Assays of Wild Type Soybean Plants

Infection assays with *P. sojae* on wild type soybean plants were performed by inoculating fresh mycelial plugs (0.5 cm diameter) on the abaxial side of detached leaves of 10-day-old soybean plants (*Glycine max* cv Opaline). Mock infections included inoculation with blank V8-agar plugs. The petioles of the detached leaves were wrapped in cotton wool and the inoculated plants were placed in a tray containing three layers of wetted absorbent paper and closed with plastic wrap foil to maintain a relative humidity of 100%. Treatments and controls were incubated in a growth room at 26°C with a 16/8 h light/dark photoperiod. Samples were collected 1, 3, and 5 days post-infection and leaves of three individual plants per treatment were pooled at each time point. Three individual biological replicates were performed.

Insect Maintenance and Non-choice Experiment with Wild Type Soybean

Aphis glycines (soybean aphid) was kindly provided by dr. Annie-Eve Gagnon (CÉROM, Quebec, Canada) and reared on soybean plants under standard conditions in a growth incubator (MLR-352 incubator, Sanyo/Panasonic, Osaka, Japan) at 25°C , 60% relative humidity and a 16 h photoperiod. In a non-choice experiment, the first trifoliate leaves of 14-day-old soybean plants were placed in a cage (Novolab) with 60 apterous adult aphids. Control samples included the cage without aphids. Three leaves from individual plants of treated and control plants were harvested and pooled after the designated time points (3, 5, and 7 days), and snap frozen in liquid nitrogen. Three individual biological replicates were performed.

Real-Time Quantitative RT-PCR

For gene expression analysis, all collected leaf and root samples were ground in liquid nitrogen and stored at -80°C until further analysis. RNA extraction was performed using TriReagent® (Sigma-Aldrich). Next, a DNase I treatment (Life Technologies) was performed and the RNA concentration and quality was assessed spectrophotometrically. First-strand cDNA was synthesized from $1\mu\text{g}$ of total RNA with oligo(dT)25 primers and 200 U of M-MLV reverse transcriptase (Life Technologies). Subsequently, the cDNA was diluted 2.5 times and cDNA quality was checked by RT-PCR with SKP1/Ask-interacting protein 16 primers (SKIP16). Quantitative RT-PCR was performed with the 96-well CFX Connect™ Real-Time PCR Detection System (Bio-Rad) using the SensiMix™ SYBR® No-ROX One-Step kit (Bioline Reagents Limited, London, UK). Reactions were conducted in a total volume of $20\mu\text{l}$ containing

$1 \times$ SensiMix™ SYBR® No-ROX One-Step mix, 500 nM gene specific forward and reverse primer and $2\mu\text{l}$ cDNA template. RT-qPCR was performed under following conditions: 10 min at 95°C , 45 cycles of 15 s at 95°C , 25 s at 60°C , and 20 s at 72°C and a melting curve was generated after every RT-qPCR run. Independent biological replicates and technical replicates were analyzed together using the sample maximization approach (Hellemans et al., 2007). An overview of all primers used in the qPCR analyses can be found in Supplementary Table 3 and the reference genes for each experiment are listed in Supplementary Table 4. Based on the available literature, different reference genes were selected because they were demonstrated to be the most stable under certain conditions. Melting curve analysis was performed after each run (Bio-Rad CFX Manager 3.1 software). Reference gene stability and quality control of the samples were validated in the qBASEPLUS software (Hellemans et al., 2007) and the results were statistically evaluated with the REST-384 software using the pair wise fixed reallocation randomization test (with 2000 randomizations; Pfaffl et al., 2002). Gene specific primers were designed using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and the specificity (BLAST search) and presence of SNPs were analyzed *in silico*, next to the secondary structure evaluation of the amplicon (Derveaux et al., 2010). Gene specific primers were evaluated by verification of the amplicon and determination of the amplification efficiency.

Germination Assays

For the seed germination assay, seeds of wild type plants and four independent homozygous transgenic lines for each construct (35S::GmNLL1 and 35S::GmNLL2) were grown on $\frac{1}{2}$ MS medium (Duchefa Biochemie, Haarlem, The Netherlands) containing 50 or 150 mM NaCl (50 seeds/line/treatment). After the stratification for 3 days at 4°C in the dark, the plates were placed in a plant growth room at 21°C and a 16/8 h light/dark cycle. Germination was assigned as the emergence of the radicle through the seed coat. Germination on $\frac{1}{2}$ MS medium without additional NaCl was performed as a control. Two biological replicates were performed with 50 plants per line for each treatment.

To determine post-germination growth, plants were sown on $\frac{1}{2}$ MS medium and after the stratification (3 days at 4°C in the dark), the plants were grown at 21°C in a plant growth room with a 16/8 h light/dark cycle. Seven-day-old plantlets were transferred to $\frac{1}{2}$ MS medium with 50 or 150 mM NaCl and after 1 week, the percentage of discolored leaves was determined. Chlorophyll was extracted by adding 10 ml N,N-dimethylformamide to the leaf material and after a 2 h incubation, the absorbance of the supernatant was measured at 645 and 663 nm. Chlorophyll *a* and *b* were determined as described by Porra (2002): $[\text{Chl } a] = 12 A_{663} - 3.11 A_{645}$, $[\text{Chl } b] = 20.78 A_{663} - 4.88 A_{645}$, and $[\text{Chl } a + b] = 17.67 A_{663} + 7.12 A_{645}$. Two biological replicates were performed with 50 plants per line for each treatment.

Root Growth Analysis

The root growth assay was performed as follows: 30 seeds of wild type plants and the different overexpression lines were

germinated on $\frac{1}{2}$ MS medium supplemented with 0, 50, or 150 mM NaCl. Plates were kept in the dark for 3 days at 4°C to break seed dormancy and were then transferred to a plant growth room at 21°C and long day (16/8 h light/dark) growth conditions. Primary root length of 2-week-old plantlets was determined with Root Detection 0.1.2 (<http://www.labutils.de/rd.html>). The experiment was repeated twice.

Non-choice Aphid Experiment with *Arabidopsis*

A permanent colony of the green peach aphid (*Myzus persicae*) was kept on sweet pepper plants under standard lab conditions (Shahidi-Noghabi et al., 2009). In a non-choice infection assay, five adult aphids were collected from rearing plants and placed on 4-week-old *Arabidopsis* leaves with a brush. After 4 days, all adult aphids were removed from the plants and the plants were returned to the plant growth incubator. On day 8, the plants were harvested and the number of nymphs and aphids residing on each plant was counted. This experiment was repeated twice with six individual plants of each line in each of the experiments.

Phytophthora Infection Assay of *Arabidopsis*

Adult rosette leaves from 4-week-old *Arabidopsis* plants were drop inoculated with 20 μ l *P. brassicae* zoospore solution (10^5 spores/ml) or mock inoculated with water. The zoospore solution was initiated as described by Bouwmeester and Govers (2009). Upon inoculation, the plants were kept in the growth cabinet under 100 % relative humidity. Samples were taken at 1, 3, 5, and 10 dpi.

Plant inoculation with pathogen mycelia was performed by placing fresh mycelium agar plugs (0.5 cm diameter) onto $\frac{1}{2}$ MS agar plates without sugar. Two-week-old *in vitro* grown *Arabidopsis* plants were placed next to the pathogen and susceptibility was evaluated 14 days post-inoculation. Mock inoculations were performed with clean V8-agar plugs.

Pseudomonas Syringae Infection Assay of *Arabidopsis*

Pseudomonas infection assays with transgenic *Arabidopsis* plants were performed as described previously with some modifications (Pieterse et al., 1996; Katagiri et al., 2002). Four-week-old *Arabidopsis* plants were spray-inoculated with the *Pseudomonas* suspension (1.6×10^7 CFU/ml in 10 mM MgSO_4 and 0.05% Silwet-L77) or mock inoculated with 10 mM MgSO_4 and 0.05% Silwet-L77. During the first 72 h after inoculation, plants were kept in 100 % relative humidity in a Conviron plant growth cabinet (Berlin, Germany). Leaves of three individual plants were sampled at 1, 2, 3, 4, and 5 dpi. Two biological replicates were performed. To estimate the lesion area, leaves were scanned with a flatbed scanner at the highest resolution. Lesion areas of individual leaves were determined in the Image Analysis Software for Plant Disease Quantification Assess 2.0 (APS, St. Paul, USA) using a self-written macro.

Arabidopsis leaves inoculated with *P. syringae* collected at 3 and 4 dpi were used for genomic DNA extraction. DNA from

approximately 100 mg of plant material was extracted using a CTAB buffer (2% CTAB, 0.1 M Tris/HCl pH 7.5; 1.4 M NaCl; 2 mM EDTA), followed by a chloroform:isoamyl alcohol (24:1) extraction. DNA was precipitated with 100% isopropanol and washed with 76% EtOH/0.2 M NaOCl and 76% EtOH/10 mM NH_4OAc . The *oprF* primers were used to target the outer membrane porin protein F gene of *P. syringae* (Brouwer et al., 2003) and *Act2* and *PEX4* primers were used as endogenous controls for *Arabidopsis* (Supplementary Table 3). The ratio of *P. syringae* genomic DNA to *Arabidopsis* DNA was calculated using REST-384 software (Pfaffl et al., 2002). Two biological replicates with two technical replicates were analyzed.

Confocal Microscopy and Image Analysis

Images were acquired with a Nikon A1R confocal laser scanning microscope (Nikon Instruments) mounted on a Nikon Ti-E inverted epifluorescence body with an S Plan Fluor ELWD 40 \times Ph2 ADM objective (NA 0.60). Different fluorescent images were acquired along the z-axis to create a picture of the complete cell. EGFP was excited with a 488 nm argon ion laser and a 515–530 nm emission filter was used. Image analysis was conducted in Fiji (Schindelin et al., 2012) and the JaCoP tool (Bolte and Cordelières, 2006) was used for colocalization analysis.

Online Tools

Prediction of protein subcellular localization and signal peptide were performed with the TargetP 1.1 and SignalP 4.1 server, respectively (Emanuelsson et al., 2000; Petersen et al., 2011). BLAST searches were conducted on the Phytozome website (<https://phytozome.jgi.doe.gov/pz/>) using default settings. Multiple sequence alignments and pairwise sequence alignments were performed with ClustalO 1.2.1 (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and EMBOSS Water (http://www.ebi.ac.uk/Tools/psa/emboss_water/), respectively. Normalized RNA-sequencing data was downloaded on the SoyBase website (<http://soybase.org/soyseq/>) (Severin et al., 2010).

Statistical Analysis

Statistical analysis was conducted using SPSS Statistics 22 (IBM) and the data were considered statistically significant for $p < 0.05$. The assumption of normality was tested with the Shapiro-Wilkinson test and the equality of variances of normally distributed data was assessed using the Levene's test. The Welch and Brown-Forsythe tests were performed when the homogeneity of variance of the data was invalid. ANOVA was used to determine statistically significant differences between groups with normally distributed data. For not-normally distributed samples, the Mann-Whitney *U*-test was performed, supplemented with the non-parametric Levene's equivalent to test homogeneity of variance. Tukey was used as *post-hoc* test with Bonferroni-Holm correction for multiple testing. This correction was also applied for Mann-Whitney tests between different groups. Data with a binomial distribution were subjected to Pearson's chi-square test. All results are shown as the mean \pm SE (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

RESULTS

The *Nictaba*-Like Lectins from Soybean Show High Sequence Similarity to *Nictaba*

In a previous study 31 genes with homology to the *Nictaba* gene from tobacco have been identified in the soybean genome (Van Holle and Van Damme, 2015). Six of them are composed of one or more *Nictaba* domains, and two of these genes, designated as *GmNLL1* and *GmNLL2*, were selected for further study. Sequence comparison between the amino acid sequences from *Nictaba* (encoded by AF389848) and the two *Nictaba*-like proteins from soybean showed that these sequences are highly related. In contrast to the tobacco lectin sequence, which only consists of a *Nictaba* domain, the *Nictaba* domain from *GmNLL1* is preceded by an N-terminal domain of 24 amino acids. The *GmNLL2* sequence encodes an N-terminal domain of 66 amino acids followed by two *Nictaba* domains separated by a 51 amino acid linker (Figure 1A). BLASTp searches revealed that the N-terminal sequences of *NLL1* and *NLL2* show no sequence homology to any other plant protein.

Amino acid sequence alignment of *Nictaba* with the *Nictaba* domains of the *GmNLLs* revealed 26 and 39% sequence identity, and 39 and 48 % sequence similarity for *NLL1* and *NLL2*, respectively. Additionally, the two Trp residues which are imperative for the carbohydrate-binding activity of the tobacco lectin (Schoupe et al., 2010), are conserved in the soybean *Nictaba* homologs (Figure 1B). The putative nuclear localization signal sequence (¹⁰²KKKK¹⁰⁵) present in the *Nictaba* sequence was not conserved in the *GmNLL* sequences (Figure 1B).

The *Nictaba*-Like Lectins from Soybean Localize to the Nucleus and Cytoplasm

Analysis of the *GmNLL* sequences using the SignalP 4.1 server (Petersen et al., 2011) indicated the absence of a signal peptide, suggesting that these proteins are synthesized on free ribosomes and reside in the cytoplasm. Since the TargetP 1.1 software (Emanuelsson et al., 2000) did not allow a clear prediction of the subcellular localization for the *GmNLLs*, fusion constructs of the *GmNLL* coding sequences N- or C-terminally tagged with EGFP were used for transient expression in *N. benthamiana* leaves. Confocal microscopy of leaf tissue at day 2 post-infiltration revealed fluorescence in the nucleus and the cytoplasm of the epidermal cells, with similar images for the N- and C-terminal EGFP fusion constructs for *NLL1* and *NLL2* (Figure 2A). Similar localization patterns were obtained after stable transformation of tobacco BY-2 suspension cells confirming that *GmNLL1* and *GmNLL2* localize to the nucleus and the cytoplasm (Figure 2B).

Expression of *NLL* Genes during Soybean Development

To investigate the expression level of the *NLL* genes in different tissues from soybean, plants were grown under normal growth conditions and different tissue samples were taken from day four after sowing until maturity of the seed pods. Transcript levels for *GmNLLs* and some classical lectins of the legume lectin

family were quantified using RT-qPCR and the expression was compared between different tissues (Figure 3).

The transcript levels for the *NLL1* gene are the highest in the cotyledons, unifoliate, and trifoliate leaves, but are significantly lower in belowground and reproductive tissues. The expression profile for the *NLL2* gene resembles that of *NLL1* with high expression in the leaves and significantly lower expression in roots. Yet, the *NLL2* transcript levels in green pods and immature seeds are higher compared to the transcript level of roots at day 4. Based on the raw Cq values of the different genes in the different samples, the expression level of the *NLL1* gene corresponds well to the expression level of the three reference genes while transcript levels for *NLL2* are less abundant than the *NLL1* gene and the reference genes (Supplementary Table 5).

The RT-qPCR analysis for the *NLL1* and *NLL2* genes was complemented with a comparative analysis to the *SVL* (soybean vegetative lectin) and *SBA* (soybean agglutinin) genes, two previously identified legume lectin genes from soybean (Figure 4). The transcript levels for *SVL* are the highest in leaves but lower transcript levels were also detected in green pods, immature seeds and roots of 19-day-old plants. In contrast very high transcript levels for the *SBA* gene were observed in pods and seeds. The expression is higher in green pods and immature seeds, compared to mature seeds. Considerably lower transcript levels of the *SBA* gene were detected in young cotyledons and in 19-day-old roots.

Nictaba-Like Genes Are Stress Inducible in Soybean

The expression patterns of *GmNLL1* and *GmNLL2* were investigated in shoots and roots of 14-day-old plants subjected to different stress treatments. The RT-qPCR data reveal that salt treatment, *P. sojae* infection and *A. glycines* infestation trigger the expression of particular *NLL* genes (Figure 5). Interestingly, the expression of the two *GmNLLs* displayed dissimilar patterns under each of the different stress treatments. Salt stress conditions triggered the transcription of the *NLL1* gene in leaves and roots (Figures 5A,B). Transcript levels in both leaves and roots reached a peak 10 h after the start of the treatment. Gene expression levels of *NLL2* in leaves and roots were not altered by salt treatment. Infection with *P. sojae* (Figure 5C) triggered both *GmNLL1* and *GmNLL2* gene expression. The upregulation of *GmNLL1* and *GmNLL2* was the highest at 3 days post-infection, being approximately 11 and 3-fold higher than the non-treated plants for *NLL1* and *NLL2*, respectively. After aphid infestation, the expression of *NLL1* and *NLL2* showed an upregulation at 5 and/or 7 days post-infection. Compared to the expression level of *NLL1*, *NLL2* was triggered to a lower extent (Figure 5D). Application of the hormones ABA and MeJA did not greatly influence the transcript levels for *GmNLL1* or *GmNLL2*. During SA treatment, the relative expression levels of *GmNLL1* and *GmNLL2* in root tissues were decreased significantly, suggesting that these gene products are not required in the plant's response upon SA treatment. The transcript levels of *GmNLL1* and *GmNLL2* in leaf tissues were not impacted by treatment with SA (Supplementary Figure 1).

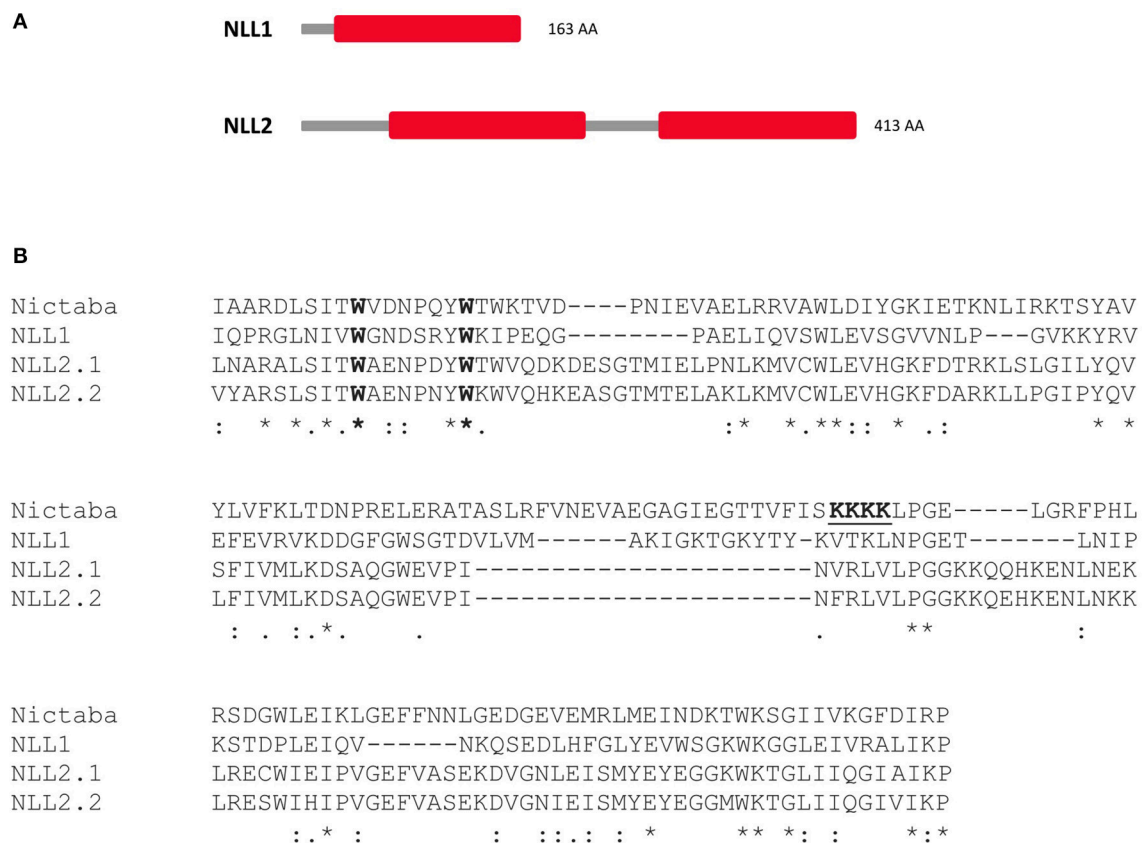


FIGURE 1 | (A) Domain architecture of the Nictaba-like homologs from soybean under study: NLL1 (*Glyma.06G221100*) and NLL2 (*Glyma.20G020900*) **(B)**

Sequence alignment of the trimmed Nictaba sequence and the Nictaba domains of NLL1 and NLL2 (NLL2.1: domain one, NLL2.2: domain two) from soybean using ClustalO. The conserved Trp-residues important for the carbohydrate-binding activity of Nictaba are marked in bold and the proposed nuclear localization signal of Nictaba is underlined. "*" indicates fully conserved amino acid residue; ":" designates conserved amino acid substitution (indicating conservation between groups of strongly similar properties); "." designates semi-conserved amino acid substitution (indicating conservation between groups of weakly similar properties).

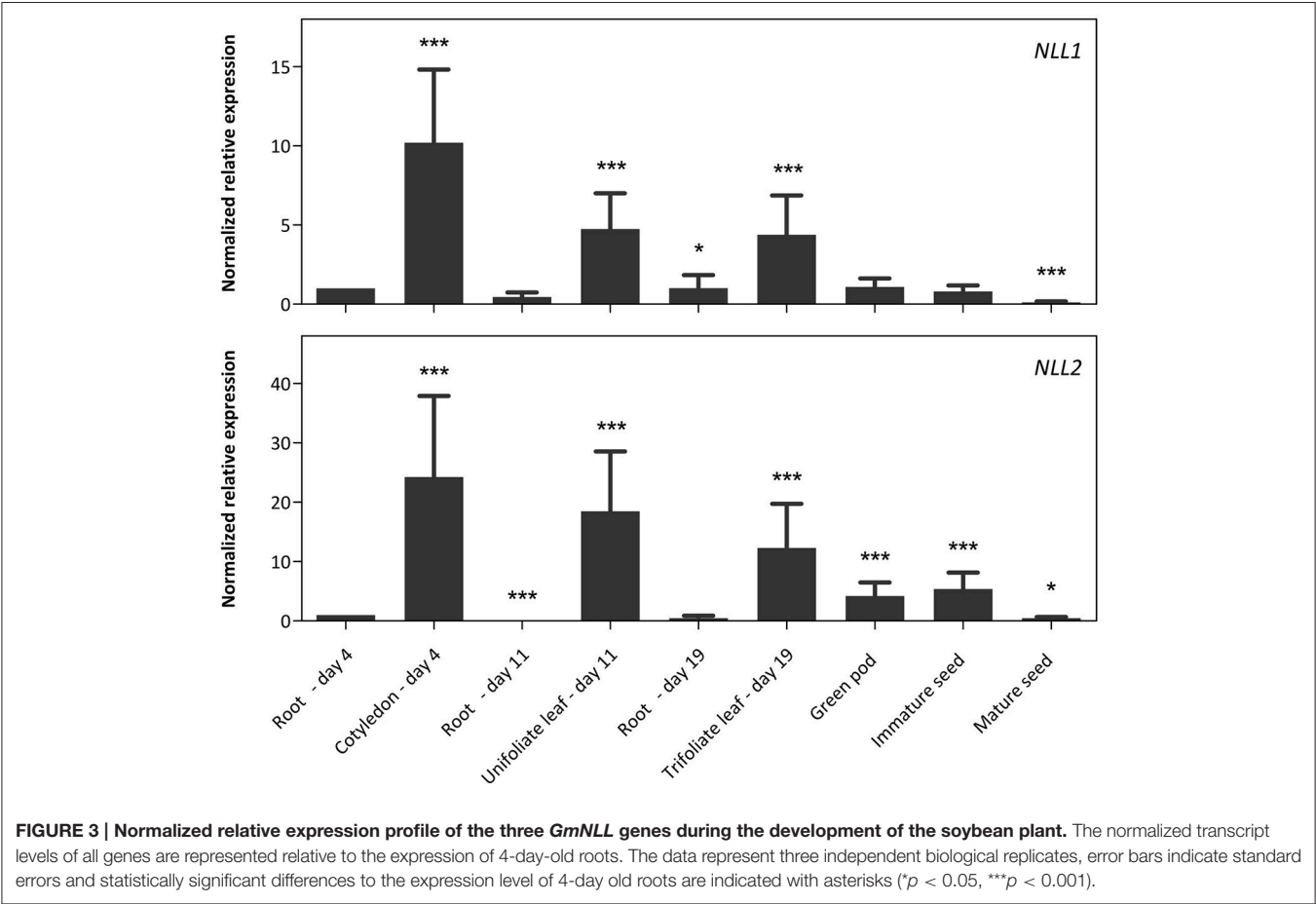
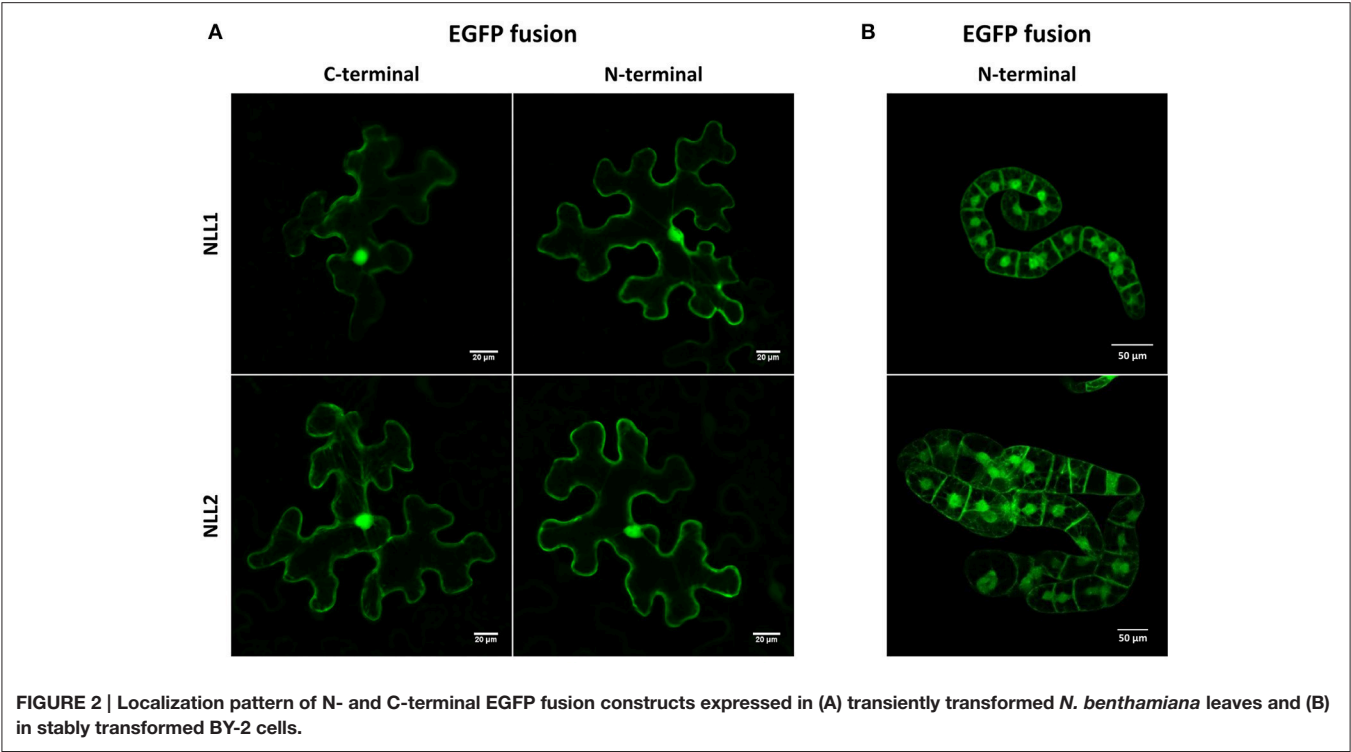
Our data show a differential expression pattern for the two *NLL* genes in both shoot and/or root tissues upon application of biotic or abiotic stresses, suggesting that these genes might play distinct roles in the plant.

Overexpression of *GmNLL1* and *GmNLL2* in *Arabidopsis* Confers Increased Tolerance to Salt Stress

To further investigate the biological function of the *GmNLLs*, transgenic *Arabidopsis* lines that overexpress *GmNLL1* or *GmNLL2* driven by the CaMV 35S promoter were generated. Several independent homozygous lines carrying a single copy of the T-DNA insertion were screened and transcript levels for *GmNLL1* and *GmNLL2* were determined by RT-qPCR in 4-week-old plants. The transcript levels relative to the expression of *TIP41* (tonoplast intrinsic protein 41), a reference gene from *Arabidopsis*, indicated that the different lines exhibited varying expression levels for the *Nictaba*-like genes. Based on these results four transgenic lines for each *GmNLL* were selected for detailed analyses (Figure 6). It should be noted that the 35S::*NLL1* lines

showed a significantly higher relative expression to *TIP41*, when compared to the 35S::*NLL2* lines.

The salt-induced expression of *GmNLL1* in soybean led us to hypothesize that *GmNLL1* might be involved in the salt stress response. In a first experiment the transgenic *Arabidopsis* lines overexpressing *GmNLL1* and *GmNLL2* were investigated for their salt stress tolerance during germination and seedling stages. Control experiments in which the germination percentage of the seeds was examined on half strength MS medium containing no salt, demonstrated that except for *NLL1-3* and *NLL2-4*, all lines exhibited the same germination percentage. Seed germination on medium containing 50 mM NaCl revealed no differences between the wild type and transgenic lines after 6 days (data not shown). On the contrary, all overexpression lines except for *NLL1-3* exhibited a similar or significantly higher germination rate on MS medium containing 150 mM NaCl compared to the wild type (Figure 7A). The lower germination percentage for *NLL1-3* and *NLL2-4* on half strength MS medium in the absence of salt could explain the lower (*NLL1-3*) or similar (*NLL2-4*) germination percentage on medium containing 150 mM NaCl.



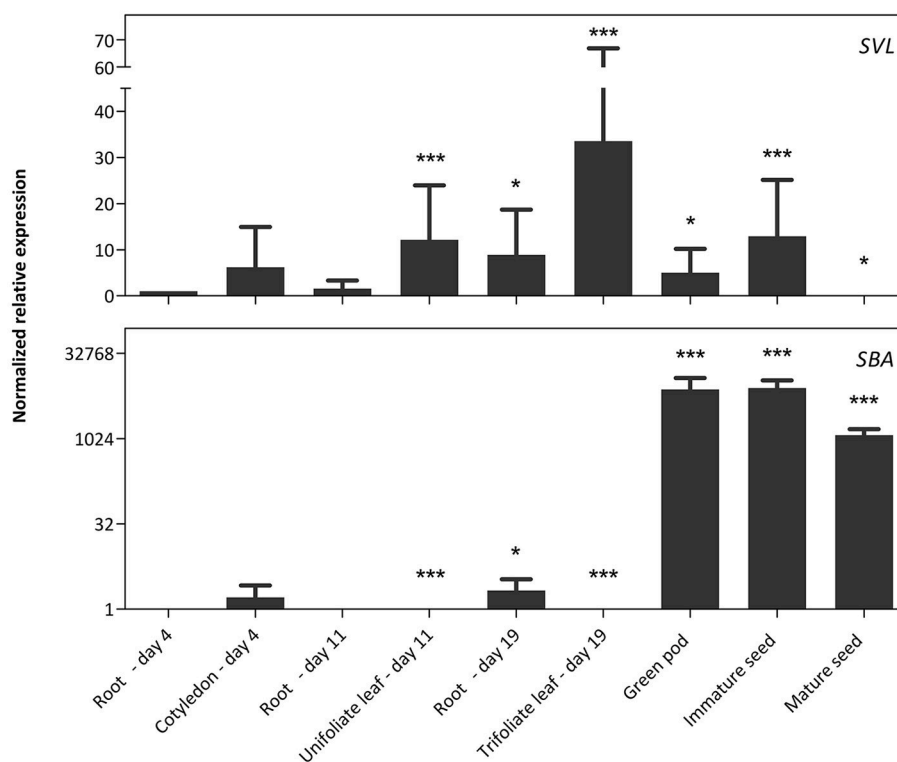


FIGURE 4 | Normalized relative expression profile of *SVL* and *SBA* during soybean development. The normalized transcript levels of all genes are represented relative to the expression of 4-day-old roots. The data represent three independent biological replicates, error bars indicate standard errors and statistically significant differences to the expression level of 4-day-old roots are indicated with asterisks (* $p < 0.05$, *** $p < 0.001$).

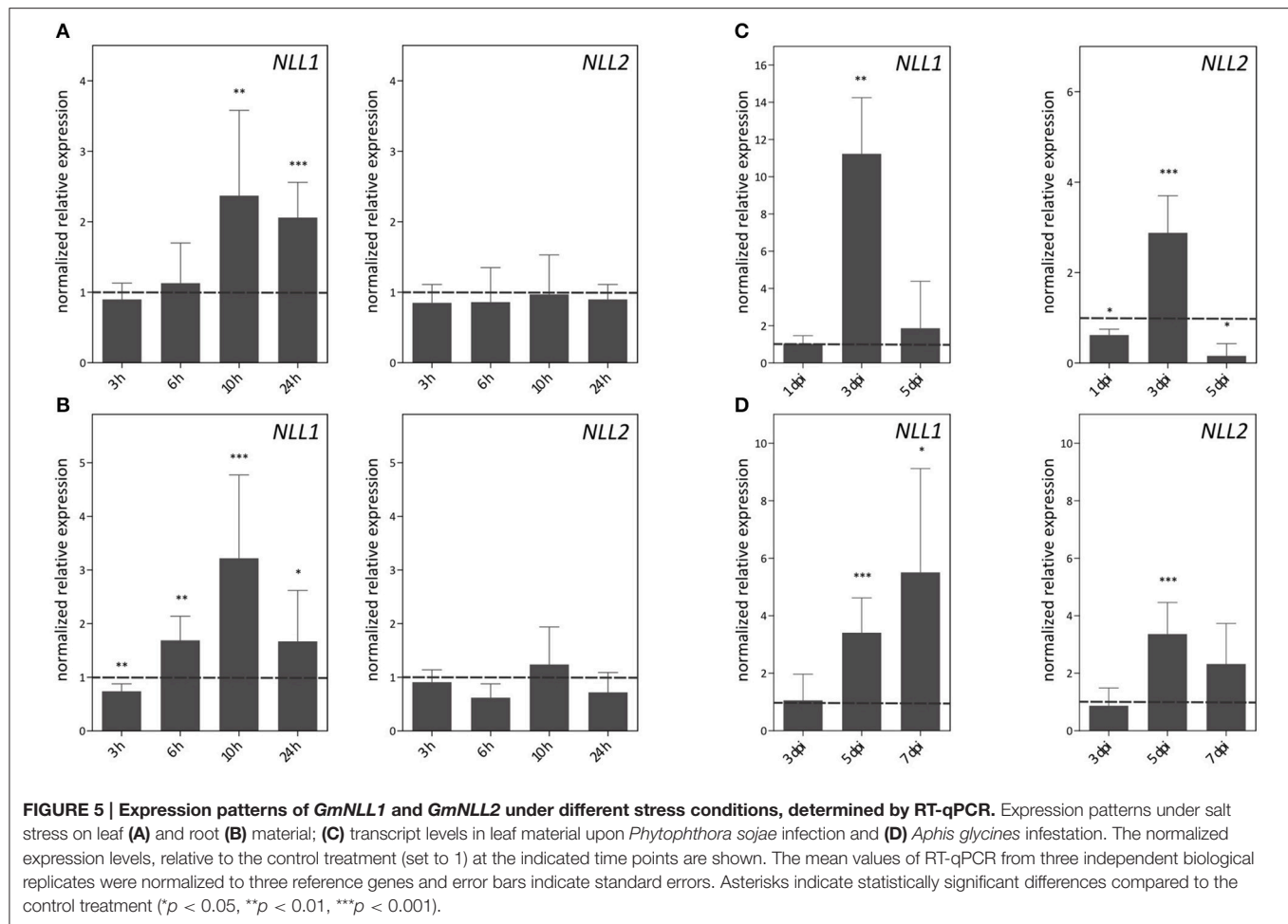
In order to explore the effect of salt stress at the seedling stage, a second experiment was performed in which the post-germination growth was investigated. The transgenic lines overexpressing *GmNLL1* and *GmNLL2* were allowed to germinate and grow on half strength MS for 1 week, and were then transferred to half strength MS supplemented with 50 or 150 mM salt. Seven days after transfer, leaf material was harvested and chlorophyll *a* and *b* were determined to estimate leaf discoloration. Under 50 mM salt conditions, no differences in chlorophyll content could be observed between wild type and transgenic plants. However, the total chlorophyll content was significantly lower for all stress treated plants compared to those of plants that had grown on normal half-strength MS medium (data not shown). When transgenic and wild type plants were transferred to medium containing 150 mM salt, the total chlorophyll content differed significantly for some of the overexpression lines (*NLL1-1*, *NLL2-1*, and *NLL2-3*) when compared to the wild type plants (Figure 7B).

In a third experiment the effect of *GmNLL1* and *GmNLL2* expression on primary root length was examined for transgenic lines and wild type plants grown in the presence of different concentrations of NaCl (0, 50, or 150 mM). No differences in primary root length were observed between wild type plants and overexpression lines grown on the normal MS medium for 14 days, nor on MS medium supplemented with 50 mM salt. However, the primary root length of transgenic lines was

significantly longer than the roots of wild type plants when plants were grown on MS supplemented with 150 mM salt (Figure 7C), suggesting that some of the *GmNLL1* and *GmNLL2* overexpression lines are more tolerant to high salt stress (150 mM NaCl) compared to wild type plants, both at the germination and the post-germination stage.

Responsiveness of the *Arabidopsis GmNLL* Overexpression Lines toward Aphids

To confirm the role of *GmNLL* in the plant defense against aphids, transgenic lines and wild type plants were infected with *M. persicae*. The observations from the two biological experiments were reproducible and the first detrimental effect of the overexpression of *GmNLL1* and *GmNLL2* was already witnessed on day 5. All adults survived on the wild type plants, while on all overexpression lines, except for *NLL2-4*, a number of the adults had died (4.1%) or started to develop wings (7.9%), suggesting that the adults found the environment unfavorable. A clear decrease in the total number of aphids on the overexpression lines compared to the wild type plants was demonstrated after 7 days (Figure 8A). Especially fewer adults resided on all overexpression lines (Figure 8B) and for some of the overexpression lines (in particular *NLL2-1* and *NLL2-4*), there is also a significant decrease in the amount of nymphs (Figure 8C).



Ectopic Expression of *GmNLL1* and *GmNLL2* in *Arabidopsis* Results in Enhanced Protection against *Pseudomonas syringae* and Does Not Enhance Plant Resistance to *Phytophthora brassicae*

Since *GmNLL1* and *GmNLL2* gene expression in soybean was significantly upregulated upon infection with *P. sojae* (Figure 5), the hypothesis was put forward that *GmNLLs* play a role in plant defense responses. The *Arabidopsis* lines overexpressing *GmNLL1* or *GmNLL2* and wild type plants were challenged with *P. brassicae* using mycelium plugs or zoospore drop inoculation to investigate the effect of *GmNLL* overexpression on the plant's resistance to pathogen infection. However, no differences in disease progression were observed between wild type plants and the *GmNLLs* overexpression lines. All plants became heavily colonized by *P. brassicae* as confirmed by staining of callose deposition in infected leaves (Results not shown).

Wild type *Arabidopsis* plants and transgenic 35S::*GmNLL1* and 35S::*GmNLL2* plants were subjected to bacterial infection with *Pseudomonas syringae* pv. *tomato* to further investigate the

role of *GmNLLs* in plant defense. Disease symptoms, bacterial growth and cell death were monitored daily. The first 2 days after the infection, no visible signs of bacterial infection were observed. Starting from 3 days post-infection, lesions were observed on the leaves and reduced disease symptoms were clear 4 days post-infection for the overexpression lines compared to the wild plants (Figure 9A and Supplementary Figure 2). In wild type plants, around 70% of the leaf is constituted of discolored lesions caused by the pathogen infection, while for all overexpression lines, the percentage of leaf damage ranged between 16 and 42% 4 days post-infection. The lesion area of mock infected plants was also measured for all time points but the calculated lesion area was never higher than 2%.

Additionally, bacterial growth of infected wild type and transgenic plants was assessed by determination of the biomass of *Pseudomonas syringae* in the inoculated *Arabidopsis* leaves. At 3 days post-infection all mean ratios for *Pseudomonas syringae* biomass in the transgenic lines are lower than those of the wild type plants (Figure 9B), but only two transgenic lines show statistically significant differences compared to the wild type plants. At 4 days post-infection, the ratios of wild type and transgenic plants

were more alike and only line *NLL2-1* demonstrated a significantly lower *Pseudomonas* biomass than the wild type.

DISCUSSION

A Nucleocytoplasmic Localization for the *GmNLL* Proteins

The two *GmNLL* genes under study are characterized by a different domain architecture. The *GmNLL1* gene encodes a Nictaba domain preceded by an N-terminal domain with unknown function while the *GmNLL2* sequence contains an unrelated N-terminal domain followed by two tandem arrayed Nictaba domains. Similar to the Nictaba sequence from tobacco, the NLL sequences from soybean do not possess a signal peptide, and are presumably synthesized on free ribosomes in the cytosol of the plant cell (Chen et al., 2002). Microscopic analysis of EGFP fusion proteins confirmed the presence of the *GmNLLs* in the cytoplasm of the plant cell, but also showed fluorescence in the nucleus. The localization of the tobacco lectin in the nucleus was initially explained by the presence of a classical nuclear localization signal, required for traditional active nuclear import (Chen et al., 2002). The functionality of the nuclear localization signal was later confirmed by Lannoo et al. (2006) since transient expression of a lectin-EGFP construct with a mutation in the nuclear localization signal sequence changed the fluorescence pattern whereby the presence of Nictaba-EGFP was restricted to the cytoplasm. Recently, these results were questioned since new localization experiments with a mutated nuclear localization signal did not affect the nucleocytoplasmic localization of the fusion protein in stably transformed tobacco suspension cultures and stably and transiently transformed *N. benthamiana* leaves, indicating that the presumed nuclear localization signal is not required for translocation of Nictaba from the cytoplasm into the nucleus (Delporte, 2013). Unlike the Nictaba sequence the *GmNLL* sequences do not contain a classical nuclear localization signal. Furthermore, *GmNLL*-GFP fusions (approximately 47 and 75 kDa for *GmNLL1* and *GmNLL2*, respectively) are too large to allow passive diffusion into the nucleus. It should be noted that additional nuclear import pathways have been characterized, depending on different import signals and these might be involved in nuclear translocation of nucleocytoplasmic lectins (Ziemięnowicz et al., 2003; Pemberton and Paschal, 2005). Thus far, it remains unclear how the soybean NLL proteins are partially translocated from the cytosol to the nucleus, similar to the tobacco lectin and other nucleocytoplasmic lectins (Al Atalah et al., 2011; Van Hove et al., 2011; Delporte, 2013). Considering the confined localization of the *GmNLLs* in the cytoplasm and nucleus, interacting partners and networks should be identified in the same cellular compartments. At present it cannot be excluded that the expression pattern would change under stress conditions, as described before for other proteins (García et al., 2010; Moore et al., 2011). Therefore, it could be interesting to investigate the localization pattern of these proteins when the plant is triggered by stress application. Expression of the GFP-NLL fusion proteins

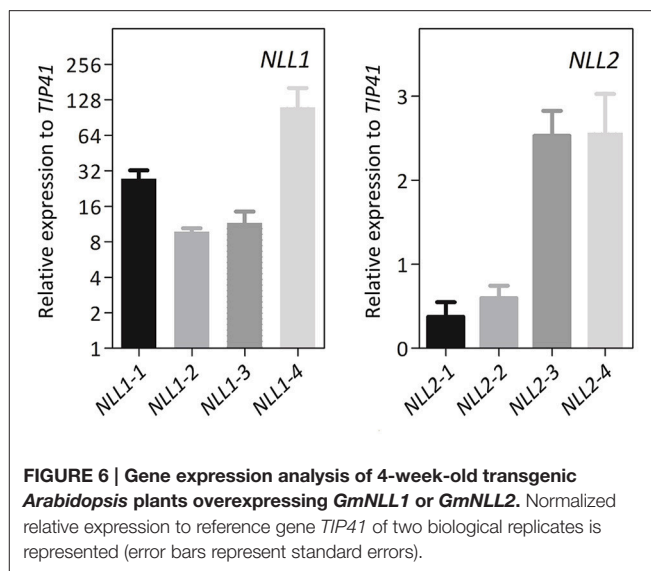


FIGURE 6 | Gene expression analysis of 4-week-old transgenic *Arabidopsis* plants overexpressing *GmNLL1* or *GmNLL2*. Normalized relative expression to reference gene *TIP41* of two biological replicates is represented (error bars represent standard errors).

under control of their own promoter could be a convenient approach.

Nictaba-Like Genes from Soybean Are Stress Inducible, Similar to the Tobacco Lectin Gene

The quantitative analysis of the *NLLs* in soybean at tissue level revealed a unique temporal and spatial expression pattern under normal environmental conditions. Although, there is high sequence similarity between the two *Nictaba*-like lectin sequences (29% sequence identity and 39% sequence similarity for the Nictaba domains), their unique expression profile suggests that a basal expression of the *NLL* genes in soybean is necessary for normal development of the soybean plant. These results are in contrast with the *Nictaba* gene from tobacco, which is not expressed under normal environmental conditions, suggesting that this protein has no role in normal growth or development of the tobacco plant (Chen et al., 2002). It was shown that only jasmonate treatment, insect herbivory and cold stress could trigger the expression of the *Nictaba* gene in tobacco (Chen et al., 2002; Vandenborre et al., 2009a, 2010; Delporte et al., 2011).

The results from our qPCR analysis are in accordance with the RNA-seq data reported by Severin et al. (2010). A comparative analysis for tissue-specific expression of the *NLL1-2* genes, the *SBA* gene, the *SVL* gene and the reference genes is represented in Supplementary Table 6. There are notable differences in the transcript levels of the root samples for the *NLL1* and *NLL2* gene. This discrepancy could be explained by differences between the developmental stages of the plant in both studies. Chragh et al. (2015) investigated the transcript levels of the *SVL* gene in 2-week-old plants by RT-qPCR and found significantly higher levels for *SVL* in unifoliate leaves compared to the other tissues analyzed. These observations are in line with our qPCR data of 11-day-old unifoliate leaf and root samples, and in agreement with the study of Saeed et al. (2008) in which the GUS

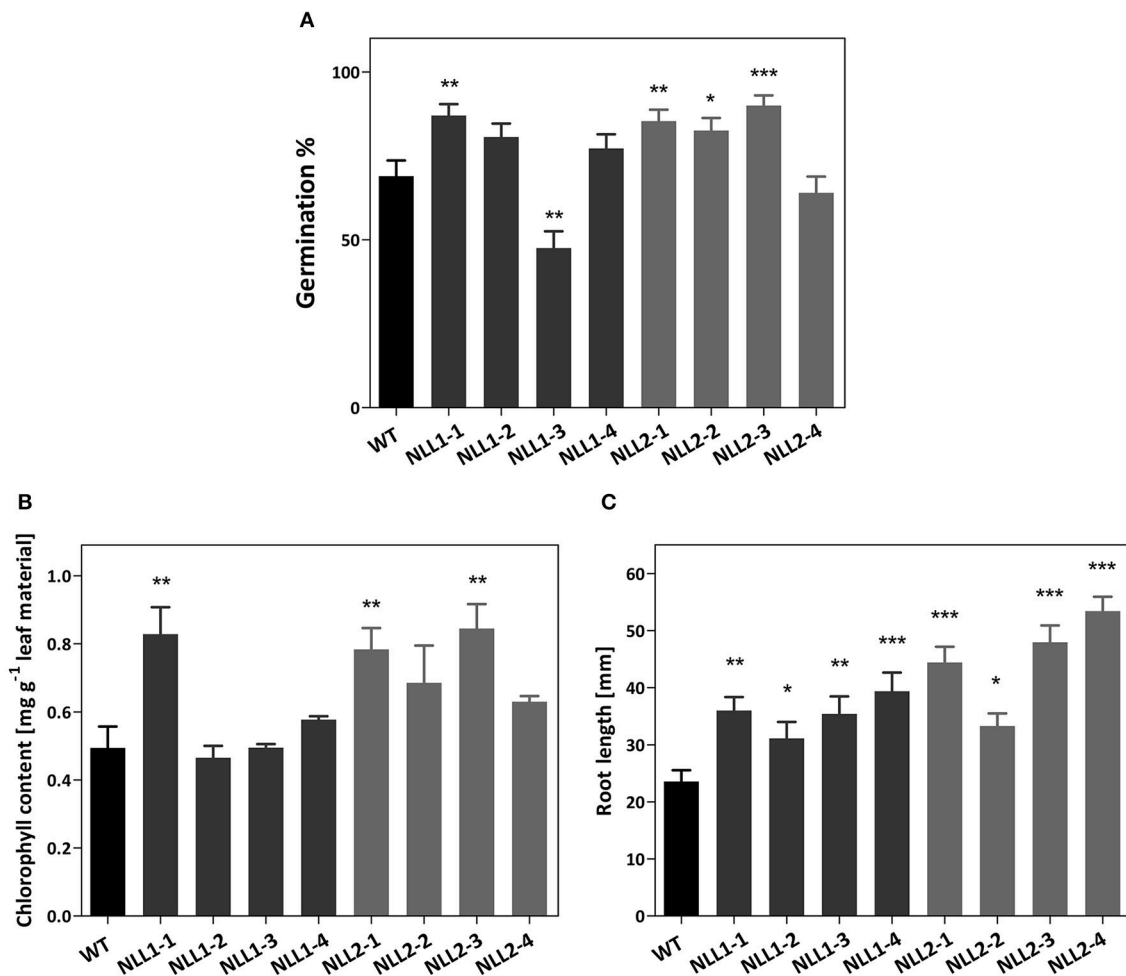


FIGURE 7 | (A) Effect of *GmNLL1* and *GmNLL2* overexpression on *Arabidopsis* germination on $\frac{1}{2}$ MS supplemented with 150 mM NaCl. Germination percentage determined on day 6, result from two biological replicates with 50 seeds per replicate, germination percentages with the corresponding standard errors are shown. **(B)** Total chlorophyll content (chlorophyll *a* + *b*) of *GmNLL1* and *GmNLL2* transgenic lines and wild type plants 7 days after transfer to $\frac{1}{2}$ MS supplemented with 150 mM NaCl. Data shows the mean \pm SE of two biological replicates. **(C)** Root length of 14-day-old *GmNLL1* and *GmNLL2* transgenic lines and wild type plants grown on $\frac{1}{2}$ MS supplemented with 150 mM NaCl. Data shows the mean \pm SE of two biological replicates with at least 30 measurements in each replicate. Asterisks indicate statistically significant differences compared to wild type (* p < 0.05, ** p < 0.01, *** p < 0.001).

reporter system was used to characterize the temporal and spatial expression of the SVL promoter in *Arabidopsis*.

Investigation of stress inducibility of the *NLL* genes demonstrated that the expression of the two *Nictaba*-like genes was induced by salt treatment (Figure 5) whereas only minor changes in *NLL* transcript levels were observed after treatment with MeJA, ABA, or SA (Supplementary Figure 1). Unexpectedly, methyl jasmonate had no effect on the expression of any of the tested *NLLs* in soybean while MeJA is one of the major triggers for the expression of *Nictaba* in tobacco (Chen et al., 2002).

Treatment with *P. sojae*, an economically important soybean pathogen, resulted in an upregulation of *GmNLL1* and *GmNLL2* (Figure 5C). These results are in agreement with the identified ESTs for *NLL1* in a cDNA isolated from *P. sojae*-infected hypocotyls (2 days post-infection; Torto-Alalibo et al., 2007).

It was demonstrated that transcript levels of *GmPR10*, one of the soybean pathogenesis-related protein genes, were already upregulated 3 h post-infection (Xu et al., 2014), indicating that *NLLs* are relatively late *P. sojae*-responsive genes. Recently, several studies focused on the elucidation of the different hormone pathways that are associated with compatible and incompatible soybean-*Phytophthora sojae* interaction. At the transcriptional level, induction of the jasmonic acid pathway was shown to be involved in compatible interactions together with suppression of the ethylene pathway and no significant changes in the SA pathway were observed (Lin et al., 2014). However, recent proteomic data revealed that different components of the SA pathway were downregulated upon infection with virulent *P. sojae* (Jing et al., 2015). The specific components and their role in the complex mechanism of the soybean-*Phytophthora sojae* interaction are not completely resolved and further investigations

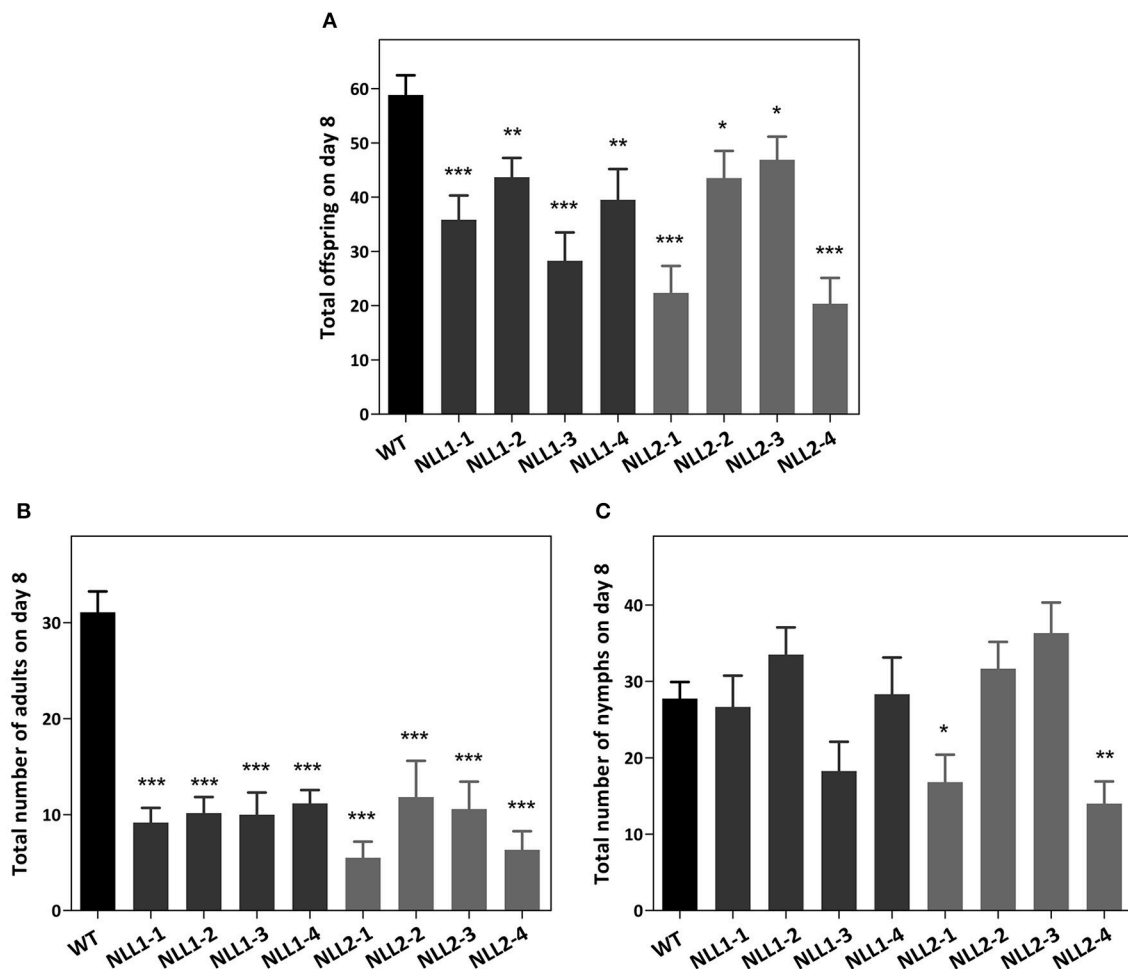


FIGURE 8 | *Myzus persicae* aphid performance in a non-choice test on wild type *Arabidopsis* plants and eight transgenic lines. The total offspring was counted after 7 days (A). The number of adults (B) and nymphs (C) residing on the plants is also shown. Values are the means \pm SE and represent the results from two biological replicates with six individual plants of every line in each of the replicates. Asterisks mark significant differences compared to wild type (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

are necessary to determine the role of the SA, ethylene and jasmonic acid pathway in this multifaceted interaction.

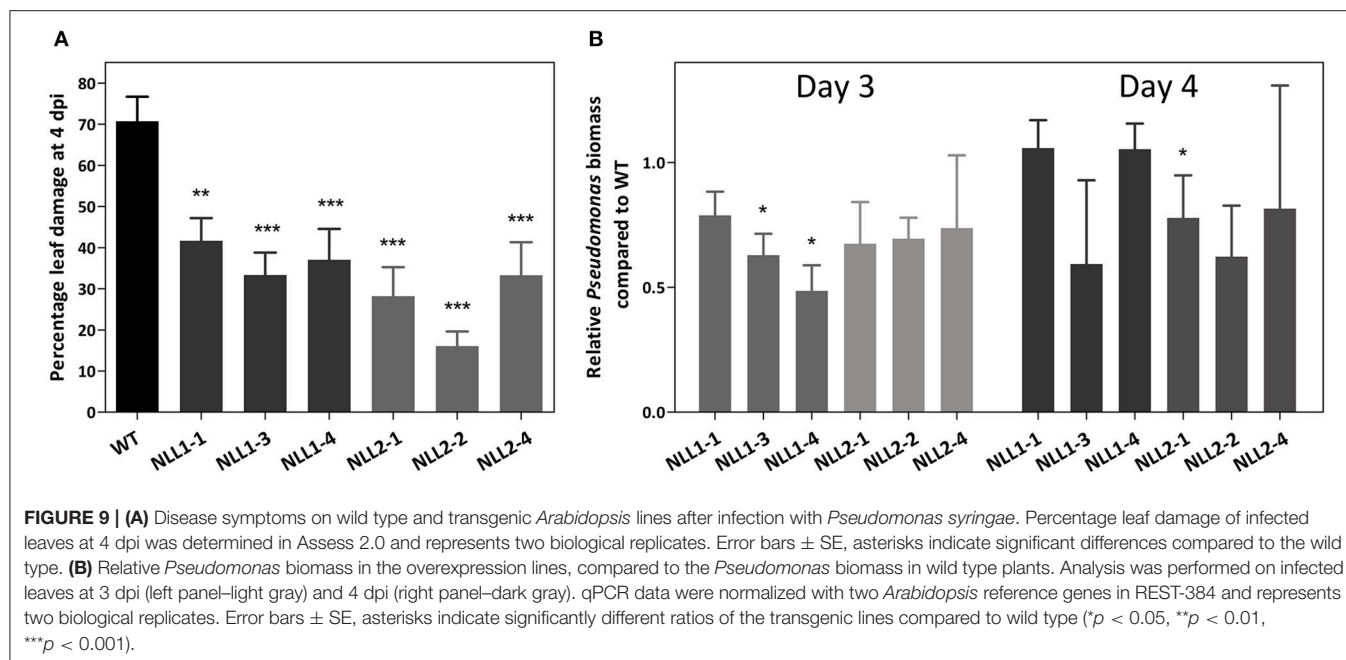
Aphis glycines infestation of soybean leaves significantly triggered the expression of *NLL1* and *NLL2*. Induction of lectin gene expression upon insect infestation was already reported for *Nictaba*. However, *Nictaba* accumulation in the tobacco plant was only upregulated after insect attack of the caterpillars *Spodoptera littoralis* and *Manduca sexta*, and the spider mite *Tetranychus urticae*. Infestation of aphids (*Myzus nicotianae*) or whiteflies (*Trialeurodes vaporariorum*) or infection with other pathogens (tobacco mosaic virus, *Botrytis cinerea* or *Pseudomonas syringae* pv. *tabaci*) did not alter the expression of the tobacco lectin (Lannoo et al., 2007; Vandenborre et al., 2009a,b).

Our results demonstrate that soybean *NLL* genes are responsive to both biotic and abiotic stresses. Such a crosstalk is orchestrated by the involvement of not only plant hormones, but also MAPK (mitogen-activated protein kinase), ROS (reactive oxygen species), transcription factors, heat shock factors and

small RNAs and was reviewed and reported for multiple plants including soybean (Fujita et al., 2006; Atkinson and Urwin, 2012; Nakashima et al., 2014; Rejeb et al., 2014; Ramegowda and Senthil-Kumar, 2015; Gupta et al., 2016).

Ectopic Expression of *GmNLLs* in *Arabidopsis* Confers Plant Tolerance to Salt Stress, Aphid Infestation and *Pseudomonas syringae* Infection

Our data show that soybean *Nictaba*-like lectins confer tolerance to salt stress in *Arabidopsis* transgenic lines. To further examine the roles of *GmNLLs* in abiotic stress tolerance, the transgenic overexpression lines and wild type plants were subjected to salt stress in multiple experimental set-ups. The data of the germination assay, post-germination assay, and root length assay indicated that overexpression of *GmNLL1* and *GmNLL2* resulted in higher tolerance to salt stress (150 mM NaCl). Nevertheless,



they do not show enhanced tolerance to mild salt (50 mM) stress conditions. Noteworthy, overexpression lines *GmNLL1-1*, *GmNLL2-1*, and *GmNLL2-3* display the highest enhanced tolerance in all salt stress related experiments. The differences between the different lines did not correlate with the expression level of the *GmNLLs* in *Arabidopsis*. It is possible that these lines have higher amounts of *GmNLLs* at the protein level but this could not be investigated since *GmNLL* specific antibodies are not available. Although, the protein abundances of the *GmNLLs* could not be determined, all overexpression lines performed better than the wild type plants in the germination and root growth experiments. The differences between the lines could be explained by a combination of post-transcriptional, translational, and degradative regulation after the expression of mRNA (Vogel and Marcotte, 2012; Feussner and Polle, 2015). Future salt stress experiments on adult *Arabidopsis* plants could be helpful to investigate whether older plants also possess these salt tolerant characteristics and if *GmNLL1* and *GmNLL2* might be components of the regulatory pathways of salt stress in plants.

Infection assays with *P. brassicae* did not show an enhanced disease resistance for the tested overexpression lines compared to wild type *Arabidopsis* plants. Bacterial blight of soybean is caused by *Pseudomonas syringae* pv. *glycinea* and can cause significant yield losses. *Arabidopsis* plants overexpressing *GmNLLs* were used in an infection assay with *Pseudomonas syringae* pv. *tomato*, an *Arabidopsis* compatible pathogen (Katagiri et al., 2002) and demonstrated that less disease symptoms were observed on the transgenic lines compared to wild type plants. These observations could be explained by reduced bacterial biomass ratios for some of the overexpression lines. It was demonstrated that *Pseudomonas syringae* induces both SA and JA pathways (Spoel et al., 2003) but RT-qPCR analysis demonstrated that these

pathways are not perturbed in the *Pseudomonas* infected *GmNLL* overexpression lines (data not shown).

Overexpression of *GmNLLs* was shown to reduce aphid performance on the transgenic *Arabidopsis thaliana* lines. Since the *GmNLLs* genes are expressed constitutively, the lectin will be present in all plant tissues and will also reach the phloem. Sucking of the phloem sap is the most likely route for the lectin to enter the aphid and interact with its tissues, metabolic processes and development. The total offspring of *M. persicae* was significantly reduced in all overexpression lines, ultimately leading to a reduced population buildup. Our results clearly showed that considerably fewer adults were present on the transgenic lines. We expect that there is a combined effect of the *GmNLLs* on survival of the aphids and in their reproduction. Future studies can focus on the mechanism(s) of the insecticidal activity. Experiments with tobacco plants indicated that *Nictaba* expression was not induced by aphid (*M. nicotianae*) feeding but insect feeding by *M. sexta*, *S. littoralis*, and *T. urticae* did trigger *Nictaba* accumulation (Lannoo et al., 2007; Vandenborre et al., 2009a,b). Furthermore, feeding experiments with transgenic tobacco plants in which the *Nictaba* gene was silenced, demonstrated that *S. littoralis* development was enhanced while overexpression of *Nictaba* led to significantly slower larval development of both *S. littoralis* and *M. sexta* (Vandenborre et al., 2010). This result confirms our hypothesis that *Nictaba*-like lectins from different species exhibit a strong direct insecticidal activity, but their specificity toward different insects apparently differs. Overexpression of the *GmNLLs* in *Arabidopsis* did not alter *PAD4* (phytoalexin deficient 4) transcript levels, a key component in the *Arabidopsis-Myzus persicae* signaling pathway (Louis and Shah, 2015; data not shown). These observations favor the role of *Nictaba*-like

proteins in defense mechanisms rather a function in signaling pathways upon insect feeding.

All previous research from NLLs focused on the model species *Arabidopsis* and tobacco. Hence, this is the first study that focusses on NLLs in a crop species. Our data show that similar to *Nictaba* in tobacco, the NLLs from soybean can also be considered as stress inducible proteins. Nevertheless, the *Nictaba*-like genes in both species act differently. The expression of *Nictaba* from tobacco is increased after treatment with jasmonates whereas this is not the case for the soybean NLLs under study. *Nictaba* expression in tobacco was enhanced after insect herbivory by caterpillars but not by aphids. For soybean, our data clearly show that *A. glycines* infestation triggers the expression of particular NLL genes. Furthermore, *GmNLL* overexpression lines in *Arabidopsis* reduced the growth and development of *M. persicae*. In addition, these transgenic lines also enhanced tolerance to salt stress at the seedling stage, and showed less disease symptoms upon *Pseudomonas syringae* infection. The data strongly suggest the involvement of *GmNLLs* in plant defense responses not only against pest or pathogens, but also in abiotic stress. These results propose that *GmNLLs* are controlled by a complex regulatory network. *GmNLL1* and *GmNLL2* are two possible candidates to further elucidate the physiological importance of the *Nictaba*-like lectins from soybean, which

can ultimately lead to novel strategies and design of crop plants with improved tolerance to changing environmental conditions.

AUTHOR CONTRIBUTIONS

SVH, EVD outlined and designed the study. SVH performed the experiments, analyzed and interpreted the data and prepared the manuscript. GS assisted with the design and interpretation of the aphid experiments. EVD conceived and supervised the experiments and critically revised the manuscript. All authors have read, revised and approved the final manuscript.

ACKNOWLEDGMENTS

We wish to thank Na Yu, Mohammed Hamshou and Weidong Li for their advice with setup and analysis of the aphid experiments. This work was supported by the Research Council of Ghent University [project 01G00515].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01590>

REFERENCES

- Al Atalah, B., De Vleeschauwer, D., Xu, J., Fouquaert, E., Höfte, M., and Van Damme, E. J. M. (2014). Transcriptional behavior of EUL-related rice lectins toward important abiotic and biotic stresses. *J. Plant Physiol.* 171, 986–992. doi: 10.1016/j.jplph.2014.04.004
- Al Atalah, B., Fouquaert, E., Vanderschaeghe, D., Proost, P., Balzarini, J., Smith, D. F., et al. (2011). Expression analysis of the nucleocytoplasmic lectin “Orysata” from rice in *Pichia pastoris*. *FEBS J.* 278, 2064–2079. doi: 10.1111/j.1742-4658.2011.08122.x
- Atkinson, N. J., and Urwin, P. E. (2012). The interaction of plant biotic and abiotic stresses: from genes to the field. *J. Exp. Bot.* 63, 3523–3544. doi: 10.1093/jxb/ers100
- Bolte, S., and Cordelières, F. P. (2006). A guided tour into subcellular colocalisation analysis in light microscopy. *J. Microsc.* 224, 13–232. doi: 10.1111/j.1365-2818.2006.01706.x
- Bouwmeester, K., and Govers, F. (2009). A novel method for efficient and abundant production of *Phytophthora brassicae* zoospores on Brussels sprout leaf discs. *BMC Plant Biol.* 9:111. doi: 10.1186/1471-2229-9-111
- Brouwer, M., Lievens, B., Van Hemelrijck, W., Van Den Ackerveken, G., Cammue, B. P. A., and Thomma, B. P. H. J. (2003). Quantification of disease progression of several microbial pathogens on *Arabidopsis thaliana* using real-time fluorescence PCR. *FEMS Microbiol. Lett.* 228, 241–248. doi: 10.1016/S0378-1097(03)00759-6
- Chen, Y., Peumans, W. J., Hause, B., Bras, J., Kumar, M., Proost, P., et al. (2002). Jasmonate methyl ester induces the synthesis of a cytoplasmic/nuclear chito oligosaccharide-binding lectin in tobacco leaves. *FASEB J.* 16, 905–907. doi: 10.1096/fj.01-0598fje
- Chragh, M., Zolotarov, Y., Saeed, H., and Strömviik, M., V (2015). Le4 is an epicotyl preferential homologue of the soybean seed-specific Le1 lectin and the vegetative Le3 lectin genes. *Plant Mol. Biol. Rep.* 33, 1779–1789. doi: 10.1007/s11105-015-0873-y
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-3113.1998.00343.x
- Delporte, A. (2013). *Expression and Nuclear Interacting Partners of the Tobacco Lectin*. Ph.D. thesis, Ghent University.
- Delporte, A., De Vos, W. H., and Van Damme, E. J. M. (2014). *In vivo* interaction between the tobacco lectin and the core histone proteins. *J. Plant Physiol.* 171, 1149–1156. doi: 10.1016/j.jplph.2014.04.008
- Delporte, A., Lannoo, N., Vandenborre, G., Ongenaert, M., and Van Damme, E. J. M. (2011). Jasmonate response of the *Nicotiana tabacum* agglutinin promoter in *Arabidopsis thaliana*. *Plant Physiol. Biochem.* 49, 843–851. doi: 10.1016/j.plaphy.2011.04.011
- Delporte, A., Van Holle, S., Lannoo, N., and Van Damme, E. J. M. (2015). The tobacco lectin, prototype of the family of *Nictaba*-related proteins. *Curr. Protein Pept. Sci.* 16, 5–16. doi: 10.2174/1389203716666150213154107
- Derveaux, S., Vandesompele, J., and Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods* 50, 227–230. doi: 10.1016/j.jymeth.2009.11.001
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Feussner, I., and Polle, A. (2015). What the transcriptome does not tell — proteomics and metabolomics are closer to the plants’ patho-phenotype. *Curr. Opin. Plant Biol.* 26, 26–31. doi: 10.1016/j.pbi.2015.05.023
- Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., et al. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* 9, 436–442. doi: 10.1016/j.pbi.2006.05.014
- García, A. V., Blanvillain-Baufumé, S., Huibers, R. P., Wiermer, M., Li, G., Gobbato, E., et al. (2010). Balanced nuclear and cytoplasmic activities of EDS1 are required for a complete plant innate immune response. *PLoS Pathog.* 6:e1000970. doi: 10.1371/journal.ppat.1000970
- Gupta, P., Sharma, R., Sharma, M. K., Sharma, M. P., Satpute, G. K., Garg, S., et al. (2016). “Signaling cross talk between biotic and abiotic stress responses in soybean,” in *Abiotic and Biotic Stresses in Soybean Production*, ed M. Miransari (New York, NY: Elsevier), 27–52.
- Harrison, S. J., Mott, E. K., Parsley, K., Aspinall, S., Gray, J. C., and Cottage, A. (2006). A rapid and robust method of identifying transformed *Arabidopsis*

- thaliana* seedlings following floral dip transformation. *Plant Methods* 2:19. doi: 10.1186/1746-4811-2-19
- Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* 8:R19. doi: 10.1186/gb-2007-8-2-r19
- Hu, R., Fan, C., Li, H., Zhang, Q., and Fu, Y.-F. (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol. Biol.* 10:93. doi: 10.1186/1471-2199-10-93
- Jing, M., Ma, H., Li, H., Guo, B., Zhang, X., Ye, W., et al. (2015). Differential regulation of defense-related proteins in soybean during compatible and incompatible interactions between *Phytophthora sojae* and soybean by comparative proteomic analysis. *Plant Cell Rep.* 34, 1263–1280. doi: 10.1007/s00299-015-1786-9
- Jones, J. D., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3
- Katagiri, F., Thilmony, R., and He, S. Y. (2002). The *Arabidopsis thaliana*-*Pseudomonas syringae* interaction. *Arab B* 1:e0039. doi: 10.1199/tab.0039
- Lannoo, N., Peumans, W. J., Van Pamel, E., Alvarez, R., Xiong, T. C., Hause, G., et al. (2006). Localization and *in vitro* binding studies suggest that the cytoplasmic/nuclear tobacco lectin can interact *in situ* with high-mannose and complex N-glycans. *FEBS Lett.* 580, 6329–6337. doi: 10.1016/j.febslet.2006.10.044
- Lannoo, N., and Van Damme, E. J. M. (2010). Nucleocytoplasmic plant lectins. *Biochim. Biophys. Acta* 1800, 190–201. doi: 10.1016/j.bbagen.2009.07.021
- Lannoo, N., and Van Damme, E. J. M. (2014). Lectin domains at the frontiers of plant defense. *Front. Plant Sci.* 5:397. doi: 10.3389/fpls.2014.00397
- Lannoo, N., Vandenborre, G., Miersch, O., Smagghe, G., Wasternack, C., Peumans, W. J., et al. (2007). The jasmonate-induced expression of the *Nicotiana tabacum* leaf lectin. *Plant Cell Physiol.* 48, 1207–1218. doi: 10.1093/pcp/pcm090
- Le, D. T., Aldrich, D. L., Valliyodan, B., Watanabe, Y., van Ha, C., Nishiyama, R., et al. (2012). Evaluation of candidate reference genes for normalization of quantitative RT-PCR in soybean tissues under various abiotic stress conditions. *PLoS ONE* 7:e46487. doi: 10.1371/journal.pone.0046487
- Libault, M., Thibivilliers, S., Bilgin, D. D., Radwan, O., Benitez, M., Clough, S. J., et al. (2008). Identification of four soybean reference genes for gene expression normalization. *Plant Genome* 1, 44–54. doi: 10.3835/plantgenome2008.02.0091
- Lin, F., Zhao, M., Baumann, D. D., Ping, J., Sun, L., Liu, Y., et al. (2014). Molecular response to the pathogen *Phytophthora sojae* among ten soybean near isogenic lines revealed by comparative transcriptomics. *BMC Genomics* 15:18. doi: 10.1186/1471-2164-15-18
- Louis, J., and Shah, J. (2015). Plant defence against aphids: the PAD4 signalling nexus. *J. Exp. Bot.* 66, 449–454. doi: 10.1093/jxb/eru454
- Moore, J. W., Loake, G. J., and Spoel, S. H. (2011). Transcription dynamics in plant immunity. *Plant Cell* 23, 2809–2820. doi: 10.1105/tpc.111.087346
- Nakashima, K., Yamaguchi-Shinozaki, K., and Shinozaki, K. (2014). The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front. Plant Sci.* 5:170. doi: 10.3389/fpls.2014.00170
- Pemberton, L. F., and Paschal, B. M. (2005). Mechanisms of receptor-mediated nuclear import and nuclear export. *Traffic* 6, 187–198. doi: 10.1111/j.1600-0854.2005.00270.x
- Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701
- Peumans, W. J., and Van Damme, E. J. M. (1995). Lectins as plant defense proteins. *Plant Physiol.* 109, 347–352. doi: 10.1104/pp.109.2.347
- Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Res.* 30:e36. doi: 10.1093/nar/30.9.e36
- Pieterse, C. M., van Wees, S. C., Hoffland, E., van Pelt, J. A., and van Loon, L. C. (1996). Systemic resistance in *Arabidopsis* induced by biocontrol bacteria is independent of salicylic acid accumulation and pathogenesis-related gene expression. *Plant Cell* 8, 1225–1237. doi: 10.1105/tpc.8.8.1225
- Porra, R. (2002). The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls *a* and *b*. *Photosyn. Res.* 73, 149–156. doi: 10.1023/A:1020470224740
- Ramegowda, V., and Senthil-Kumar, M. (2015). The interactive effects of simultaneous biotic and abiotic stresses on plants: mechanistic understanding from drought and pathogen combination. *J. Plant Physiol.* 176, 47–54. doi: 10.1016/j.jplph.2014.11.008
- Rejeb, I. B., Pastor, V., and Mauch-Mani, B. (2014). Plant responses to simultaneous biotic and abiotic stress: molecular mechanisms. *Plants* 3, 458–475. doi: 10.3390/plants3040458
- Saeed, H. A., Vodkin, L. O., and Strömvik, M., V. (2008). Promoters of the soybean seed lectin homologues Le2 and Le3 regulate gene expression in vegetative tissues in *Arabidopsis*. *Plant Sci.* 175, 868–876. doi: 10.1016/j.plantsci.2008.08.013
- Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open source platform for biological image analysis. *Nat. Methods* 9, 676–682. doi: 10.1038/nmeth.2019
- Schoupe, D., Ghesquière, B., Menschaert, G., De Vos, W. H., Bourque, S., Trooskens, G., et al. (2011). Interaction of the tobacco lectin with histone proteins. *Plant Physiol.* 155, 1091–1102. doi: 10.1104/pp.110.170134
- Schoupe, D., Rougé, P., Lasanajak, Y., Barre, A., Smith, D. F., Proost, P., et al. (2010). Mutational analysis of the carbohydrate binding activity of the tobacco lectin. *Glycoconj. J.* 27, 613–623. doi: 10.1007/s10719-010-9305-2
- Severin, A. J., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Shahidi-Noghabi, S., Van Damme, E. J. M., and Smagghe, G. (2009). Expression of *Sambucus nigra* agglutinin (SNA-I) from elderberry bark in transgenic tobacco plants results in enhanced resistance to different insect species. *Transgenic Res.* 18, 249–259. doi: 10.1007/s11248-008-9215-2
- Sparkes, I. A., Runions, J., Kearns, A., and Hawes, C. (2006). Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* 1, 2019–2025. doi: 10.1038/nprot.2006.286
- Spoel, S. H., Koornneef, A., Claessens, S. M. C., Korzelius, J. P., Van Pelt, J. A., Mueller, M. J., et al. (2003). NPR1 modulates cross-talk between salicylate- and jasmonate-dependent defense pathways through a novel function in the cytosol. *Plant Cell* 15, 760–770. doi: 10.1105/tpc.009159
- Stefanowicz, K., Lannoo, N., Proost, P., and Van Damme, E. J. M. (2012). *Arabidopsis* F-box protein containing a Nictaba-related lectin domain interacts with N-acetylglucosamine structures. *FEBS Open Bio* 2, 151–158. doi: 10.1016/j.fob.2012.06.002
- Stefanowicz, K., Lannoo, N., Zhao, Y., Eggermont, L., Van Hove, J., Al Atalah, B., et al. (2016). Glycan-binding F-box protein from *Arabidopsis thaliana* protects plants from *Pseudomonas syringae* infection. *BMC Plant Biol.* 16:213. doi: 10.1186/s12870-016-0905-2
- Torto-Alalibo, T. A., Tripathy, S., Smith, B. M., Arredondo, F. D., Zhou, L., Li, H., et al. (2007). Expressed sequence tags from *Phytophthora sojae* reveal genes specific to development and infection. *Mol. Plant Microbe Interact.* 20, 781–793. doi: 10.1094/MPMI-20-7-0781
- Van Damme, E. J. M., Lannoo, N., Fouquaert, E., and Peumans, W. J. (2004). The identification of inducible cytoplasmic/nuclear carbohydrate-binding proteins urges to develop novel concepts about the role of plant lectins. *Glycoconj. J.* 20, 449–460. doi: 10.1023/B:GLYC.0000038291.67527.a5
- Van Damme, E. J. M., Lannoo, N., and Peumans, W. J. (2008). Plant lectins. *Adv. Bot. Res.* 48, 107–209. doi: 10.1016/S0065-2296(08)00403-5
- Vandenborre, G., Groten, K., Smagghe, G., Lannoo, N., Baldwin, I. T., and Van Damme, E. J. M. (2010). *Nicotiana tabacum* agglutinin is active against Lepidopteran pest insects. *J. Exp. Bot.* 61, 1003–1014. doi: 10.1093/jxb/erp365
- Vandenborre, G., Miersch, O., Hause, B., Smagghe, G., Wasternack, C., and Van Damme, E. J. M. (2009b). *Spodoptera littoralis*-induced lectin expression in tobacco. *Plant Cell Physiol.* 50, 1142–1155. doi: 10.1093/pcp/pcp065
- Vandenborre, G., Van Damme, E. J. M., and Smagghe, G. (2009a). *Nicotiana tabacum* agglutinin expression in response to different biotic challengers. *Arthropod. Plant Interact.* 3, 193–202. doi: 10.1007/s11829-009-9075-6
- Van Holle, S., and Van Damme, E. J. M. (2015). Distribution and evolution of the lectin family in soybean (*Glycine max*). *Molecules* 20, 2868–2891. doi: 10.3390/molecules20022868

- Van Hove, J., De Jaeger, G., De Winne, N., Guisez, Y., and Van Damme, E. J. M. (2015). The *Arabidopsis* lectin EULS3 is involved in stomatal closure. *Plant Sci.* 238, 312–322. doi: 10.1016/j.plantsci.2015.07.005
- Van Hove, J., Fouquaert, E., Smith, D. F., Proost, P., and Van Damme, E. J. M. (2011). Lectin activity of the nucleocytoplasmic EUL protein from *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* 414, 101–105. doi: 10.1016/j.bbrc.2011.09.031
- Vogel, C., and Marcotte, E. M. (2012). Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13, 227–232. doi: 10.1038/nrg3185
- Xu, P., Jiang, L., Wu, J., Li, W., Fan, S., and Zhang, S. (2014). Isolation and characterization of a pathogenesis-related protein 10 gene (GmPR10) with induced expression in soybean (*Glycine max*) during infection with *Phytophthora sojae*. *Mol. Biol. Rep.* 41, 4899–4909. doi: 10.1007/s11033-014-3356-6
- Ziemienowicz, A., Haasen, D., Staiger, D., and Merkle, T. (2003). *Arabidopsis* transportin1 is the nuclear import receptor for the circadian clock-regulated RNA-binding protein AfGRP7. *Plant Mol. Biol.* 53, 201–212. doi: 10.1023/B:PLAN.0000009288.46713.1f

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Van Holle, Smagghe and Van Damme. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Understanding the Impact of Drought on Foliar and Xylem Invading Bacterial Pathogen Stress in Chickpea

Ranjita Sinha, Aarti Gupta and Muthappa Senthil-Kumar *

National Institute of Plant Genome Research, New Delhi, India

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Zhilong Bao,
University of Florida, USA
Hans-Peter Kaul,
University of Natural Resources and
Life Sciences, Vienna, Austria

*Correspondence:

Muthappa Senthil-Kumar
skmuthappa@nipgr.ac.in

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 21 April 2016

Accepted: 08 June 2016

Published: 21 June 2016

Citation:

Sinha R, Gupta A and
Senthil-Kumar M (2016)
Understanding the Impact of Drought
on Foliar and Xylem Invading Bacterial
Pathogen Stress in Chickpea.
Front. Plant Sci. 7:902.
doi: 10.3389/fpls.2016.00902

In field conditions, plants are concurrently exposed to multiple stresses, where one stressor impacts the plant's response to another stressor, and the resultant net effect of these stresses differs from individual stress response. The present study investigated the effect of drought stress on interaction of chickpea with *Pseudomonas syringae* pv. phaseolicola (Psp; foliar pathogen) and *Ralstonia solanacearum* (Rs; xylem inhabiting wilt causing pathogen), respectively, and the net-effect of combined stress on chlorophyll content and cell death. Two type of stress treatments were used to study the influence of each stress factor during combined stress, viz., imposition of drought stress followed by pathogen challenge (DP), and pathogen inoculated plants imposed with drought in course of pathogen infection (PD). Drought stress was imposed at different levels with pathogen inoculum to understand the influence of different stress intensities on stress interaction and their net impact. Drought stressed chickpea plants challenged with Psp infection (DPsp) showed reduced *in planta* bacterial number compared to Psp infection alone. Similarly, Rs infection of chickpea plants showed reduced *in planta* bacterial number under severe drought stress. Combined drought and Psp (DPsp) infected plants showed decreased cell death compared to plants infected only with Psp but the extent of cell death was similar to drought stressed plants. Similarly, chlorophyll content in plants under combined stress was similar to the individual drought stressed plants; however, the chlorophyll content was more compared to pathogen only infected plants. Under combined drought and Rs infection (DRs), cell death was similar to individual drought stress but significantly less compared to only Rs infected plants. Altogether, the study proposes that both stress interaction and net effect of combined stress could be majorly influenced by first occurring stress, for example, drought stress in DP treatment. In addition, our results indicate that the outcome of the two stress interaction in plant depends on timing of stress occurrence and nature of infecting pathogen.

Keywords: chickpea, combined stress, biotic-abiotic stress interaction, drought, *Ralstonia solanacearum*, *Pseudomonas syringae* pv. phaseolicola

INTRODUCTION

Chickpea (*Cicer arietinum*) is an important agricultural crop as well as second largest produced legume in the world (Gaur et al., 2012). However, the global productivity of chickpea is continually challenged by abiotic and biotic stresses. Chickpea plants are vulnerable to prolonged drought stress which causes around 40–50% yield loss (Gaur et al., 2012). In addition, biotic stresses including wilt (caused by *Fusarium oxysporum*) and foliar diseases such as Ascochyta blight (caused by *Ascochyta rabiei*) and botrytis gray mold (caused by *Botrytis cinerea*) have devastating effect on chickpea cultivation (Nene et al., 2012). In view of this, several studies have been pursued to understand the molecular mechanism of stress tolerance in response to individual stresses (Jha et al., 2014; Li et al., 2015); however, this knowledge could not be directly extrapolated for improving the stress tolerance against combined stresses. Plants in field conditions are continually exposed to multiple abiotic and biotic stresses, which results in altered physiological and biochemical changes and ultimately influence yield (Ramegowda and Senthil-Kumar, 2014; Suzuki et al., 2014), and therefore, investigating the impact of combined stress is imperative in plant stress biology. Studies indicate that the plants exhibit certain unique physiological and molecular responses in addition to several common responses for circumventing the combined effect of these stresses (Choi et al., 2013; Prasch and Sonnewald, 2013; Gupta et al., 2016b).

In combined stress scenario, drought can positively or negatively affect pathogen infection (Mattson and Haack, 1987). Drought stress may also influence the pathogen virulence or pathogenicity, resulting in upsurge of different potential pathogens not known earlier (Desprez-Loustau et al., 2007; Yáñez-López et al., 2012). Previous reports have shown that drought increases the susceptibility of plant to bacterial pathogens (Mohr and Cahill, 2003; Choi et al., 2013). In chickpea, drought stress has been shown to predispose the plant and significantly increase the incidence of dry root rot caused by *Rhizoctonia bataticola* (Sharma and Pande, 2013). Contrastingly, drought stress has also been shown to enhance the tolerance toward bacterial pathogen (Ramegowda et al., 2013; Gupta et al., 2016a). On the other hand, pathogens are also shown to influence plant-water relations (Mattson and Haack, 1987; Beattie, 2011). For example, pathogen can cause water soaking in infected leaf (Beattie, 2011) and vascular wilts can induce physiological drought stress on plants (Yadeta and Thomma, 2013). The two co-occurring stressors can modulate plant responses in a way different from when the two stressors occur independently. Earlier evidences suggest that the net effect of drought and bacterial pathogen combination on plant physiology and yield is different from the individual stresses. For example, *Xylella fastidiosa* (causal agent of Pierce's disease) infection in *Vitis vinifera* under drought stress showed increase in disease symptoms and decrease in leaf water potential, net CO₂ assimilation, stomatal conductance, and transpiration rate (Choi et al., 2013).

The present study was conducted in chickpea plants exposed to combined drought stress and infection with *Pseudomonas*

syringae pv. *phaseolicola* (Psp; foliar bacterial pathogen) and *Ralstonia solanacearum* (Rs; xylem inhabiting wilt causing bacterial pathogen) for testing three notions; (1) the impact of one stress on plant's interaction with other stress; (2) influence of order of stress occurrence and severity of each stress on the outcome of stress interaction; and (3) difference in the net impact of combined stress compared to two independent stresses.

Psp causes halo blight in broad bean (Saettler, 1991), a legume closely related to chickpea (Zhu et al., 2005). Halo blight appears as water soaked lesions. Rs is known to infect more than 200 plants species (Genin, 2010) including *Medicago truncatula*, another species closely related to chickpea and various other legume plants (Vailleau et al., 2007). Rs colonizes xylem tissue and secretes exopolysaccharides which inhibits the water supply of host plant which eventually results in vascular dysfunction and wilting (Genin, 2010).

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of *Cicer arietinum* varieties PUSA 372 (procured from Indian Agriculture Research Institute, New Delhi) and ICC 4958 (available in our institute) were germinated in pots (3 inch in diameter) having a mixture of air dried peat (Prakruthi Agri Cocopeat Industries, Karnataka, India) and vermiculite (3:1, vol/vol) (Keltech Energies Pvt Ltd., Maharashtra, India) in an environmentally controlled growth chamber (PGR15, Conviron, Winnipeg, Canada) with diurnal cycle of 12-h-light/12-h-dark, 200 $\mu\text{E m}^{-2}\text{s}^{-1}$ photon flux intensity, 22°C temperature and 70% relative humidity. Pots were bottom irrigated every 2 days with half strength Hoagland's medium (TS1094, Hi-media Laboratories, Mumbai, India).

Bacterial Pathogen Inoculum Preparation

Pure culture of bacterial pathogens, viz. *Pseudomonas syringae* pv. *phaseolicola* (Psp) and *Ralstonia solanacearum* (Rs, procured from Indian type culture collection BI0001), IARI, New Delhi were used in this study. A single colony of Psp was inoculated in King's B (KB) medium (M1544, Hi-media Laboratories, Mumbai, India) supplemented with rifampicin (50 $\mu\text{g/mL}$) and incubated at 28°C with a continuous shaking of 200 rpm for 12 h. Rs was inoculated in LB medium (M124, Hi-media Laboratories, Mumbai, India) (without antibiotic) and incubated at 28°C with a continuous shaking of 200 rpm for 4 h. Both Psp and Rs were grown till the optical density (OD₆₀₀) reached 0.6 and the cultures were pelleted down at 3500 g for 10 min. The pellets were washed twice with sterile distilled water and diluted to desired concentrations by re-suspending in sterile distilled water. The OD₆₀₀ = 0.005 corresponding to 7×10^5 colony forming units (cfu) /mL for Rs and 2.5×10^6 cfu/mL for Psp were used for infecting the plants. CfU corresponding to desired OD was calculated by plating the different dilutions for OD₆₀₀ = 0.005.

Pathogen Inoculation

To study the pathogenicity of bacterial strains in chickpea ICC4958 (12-d-old), Psp suspension corresponding to 2.5×10^6 cfu/mL was syringe infiltrated into the leaves and *in planta*

bacterial number was determined from 0 to 10 days post-inoculation (dpi). Rs suspension (7×10^5 cfu/mL) was vacuum infiltrated into the plants. For this, plants were placed inverted in a beaker containing Rs suspension with 0.02% Silwet L77 (Lehle seeds, Fisher Scientific, MA, USA) and vacuum of 8.7 psi was applied for 10 min. Plants were rinsed in water immediately after infiltration and *in planta* Rs number and phenotypic symptoms were recorded from 0 to 10 dpi. The leaf infiltration of Rs was previously reported in tobacco leaves (Kiba et al., 2003), and the infected plants displayed phenotypic disease symptoms similar to the symptoms observed by root inoculation method (Kanda et al., 2003; Shinohara et al., 2005). Similarly, syringe infiltration technique used for inoculation of Psp is a well-established technique (Liu et al., 2015).

Drought Imposition

Chickpea plants were grown in a pre-weighed pot mix and were subjected to drought stress by withholding the water supply. Drought stress levels were measured in terms of pot mix field capacity (FC) using gravimetric method (Reynolds, 1970), wherein for example, plants at FC 20% perceived 80% drought. Three drought levels viz. 60, 40, and 20% FC were used in the study and they were termed as mild, moderate and severe drought, respectively (Supplementary Table S1). A pot mix pre-maintained at 80% FC (with plant) took 2, 4, and 6 days to achieve 60, 40, and 20% FC, respectively, after withholding the water. In order to achieve all the drought levels on the same day, water was withheld on every alternate day for 3 batches, which resulted in generation of three sets of plants at 20, 40, and 60% FC on the sixth day. Plants with 80% FC were maintained as controls (Supplementary Figure S1). The respective FCs were maintained by adding the lost amount of water, till the end of the experiment.

Combined Stress Imposition

Two methods were used for imposing combined stress, viz. drought followed by pathogen (DP) and pathogen followed by drought (PD). For DP studies, chickpea plants (20-d-old) with 60, 40, and 20% FC were vacuum infiltrated with Psp ($OD_{600} = 0.005$; 2.5×10^6 cfu/mL) and Rs ($OD_{600} = 0.005$; 7×10^5 cfu/mL) using aforementioned protocols. The pot surfaces were sealed with cellophane tape to avoid the entry of bacterial suspension into the pot mix (which may otherwise change the FC). After inoculation, the plants were sprayed with water and surface water was removed by blotting. Plants infiltrated with water (supplemented with 0.02% Silwet L77) were treated as mock.

For PD studies, chickpea plants were vacuum infiltrated with 2.5×10^6 cfu/mL of Psp and 7×10^5 cfu/mL of Rs following above mentioned protocol. Plants infiltrated with sterile water (supplemented with 0.02% Silwet L77) were treated as mock. Drought stress was imposed on plants 1 day after bacterial infection. A batch of plants infected with Psp and Rs was maintained without drought stress treatment (pathogen only stress, 80% FC). Similarly, a batch of uninfected plants subjected to drought stress only was maintained. Absolute control plants without bacterial as well as drought treatments were maintained at 80% FC. The experimental design for combined stress imposition is summarized in Supplementary Figures S1, S2.

Sample Harvest

Chickpea leaflets were harvested from the third twig (from hypocotyl). For DP stress, three leaflets from the same leaf were collected at 0 dpt for *in planta* bacterial multiplication assay. At 2 dpt, one leaf was collected for RNA isolation and 3 leaflets from leaf was used for *in-planta* bacterial multiplication. At 6 dpt, 3 leaflets for *in-planta* bacterial multiplication and 2 leaflet for cell death were collected from the same leaf. Three leaflet sample for the total chlorophyll and 3 leaf for phenotypic assessment were collected at 12 dpt. For PD stress, leaflet samples were collected at 0, 1, 2, 3, and 10 days post infection. The technical replicates were collected from same leaf. The methodology adopted for the sample collection and other experimental details are illustrated in detail in Supplementary Figures S1–S3.

Assay for Quantification of *in Planta* Bacterial Number

The infected leaflets were surface sterilized with 0.01% H_2O_2 for 5 s, weighed and homogenized in 100 μ L of sterile water. The homogenate was serially diluted in sterile water and the dilutions were plated on KB agar medium supplemented with rifampicin and on LB medium for assaying Psp and Rs counts, respectively. Total bacterial numbers were calculated as Log (cfu/mg fresh weight of leaf; Wang et al., 2012) and Log (cfu/mg dry weight of leaf) (Supplementary Figure S9).

Bacterial number (Cfu/mg) was calculated using the formula:

$$\text{Cfu/mg} = \frac{\text{Number of colonies} \times \text{volume of homogenate} \times \text{dilution factor}}{\text{volume plated} \times \text{weight of the leaflet (mg)}}$$

Estimation of Total Chlorophyll Content

Chlorophyll content of chickpea leaf discs [12.57 mm² (4 mm diameter)] was determined at 12 dpt for DP, drought only, pathogen only, absolute control and mock control samples using method described by Hiscox and Israelstam (1980) with minor modifications. Leaves were incubated in 1 mL of dimethyl sulfoxide (DMSO): acetone (1:1 vol/vol) mix at room temperature in dark condition for 72 h for total chlorophyll extraction. Absorbance of extracts was read using Shimadzu UV 1800 spectrophotometer (Shimadzu Corporation, Kyoto, Japan) at 645 and 663 nm. Total chlorophyll content was calculated according to Arnon's equation (Arnon, 1949).

$$\begin{aligned} &\text{Total chlorophyll content } (\mu\text{g/cm}^2) \\ &= \frac{[(\text{ml solvent})(20.2 \times \text{absorbance } 645) + (8.02 \times \text{absorbance } 663)]}{\text{leaf area (mm square)}} \end{aligned}$$

Cell Death

Cell death assay was performed as described by Koch and Slusarenko (1990) with minor modifications. Leaf samples from DP, drought only, pathogen only, absolute control and mock control were immersed in lactophenol-trypan blue for 12 h at room temperature followed by overnight de-staining in chloral hydrate (500 gm dissolved in 200 mL water). Lactophenol-trypan blue was prepared by dissolving 10 mL of lactic acid, 10 mL of glycerol, 10 g of phenol

and 10 mg of trypan blue in 10 mL of distilled water. Cell death was observed under bright field microscope (Nikon Eclipse 80i, Nikon Corporation, Tokyo, Japan). The intensity of trypan blue staining was quantified using ImageJ software (<http://imagej.nih.gov/ij/>) (Schneider et al., 2012).

Real-Time PCR Analysis

Expression profiles of genes responsive to drought (*CaLEA1*, *CaLEA2*, *CaLEA4*, *CaDREB2A*, and *CaNCED1*) and pathogen (*CaPAL2* and *CaPR4*) in DP (2 dpt), PD (10 dpt) and their respective individual stressed samples were analyzed by quantitative real-time PCR (RT-qPCR) (gene list with accession number given in Supplementary Table S2). Total RNA from leaf samples (100 mg fresh weight) was isolated using TriZol reagent (Cat # 15596018, Thermo Fisher Scientific, California, USA) following manufacturer's guidelines. RNA quality was ascertained by agarose gel electrophoresis and quantified using NanoDrop ND-1000 spectrophotometer (Thermo Scientific, MA, USA). RNA samples with OD ratios in the range of 1.9–2.1 at 260/280 nm, and 2.0–2.3 at 260/230 nm were used for cDNA synthesis. First strand cDNA was synthesized using Verso cDNA synthesis kit (Cat # K1621, Thermo Fisher Scientific, MA, USA) from 5 µg of DNase treated total RNA in a reaction volume of 50 µL. The primers used in this study were synthesized from Sigma-aldrich, USA (Supplementary Table S2). Reaction mix comprised of 1 µL of 5 fold-diluted cDNA, 1 µL of each primer (10 µM/µL) and 5 µL of SYBR Green PCR master mix (Cat # 4309155, Thermo Fisher Scientific, MA, USA) in a final volume of 10 µL. The reaction was run in ABI Prism 7000 sequence detection system (Applied Biosystems, California, USA). *CaACT1* (EU529707.1) and *Ca18S* (AJ577394.1) genes were used as endogenous control, and the cycle threshold (Ct) values obtained for these genes were used to normalize the data for PD and DP experiments, respectively. Relative fold change in gene expression was quantified using $2^{-\Delta\Delta C_t}$ method (Livak and Schmittgen, 2001). Expression analysis was carried out using two independent biological replicates. For statistical analysis, the relative quantification value (RQ) was transformed to log₂ value and test of significance was performed by one sample *t*-test. Relative transcript abundance of the chosen genes in DP, PD, and pathogen only samples was normalized with mock control, and expression profile of these genes in drought only sample was compared with absolute control.

Statistical Analysis

Data represented in the present study is derived from single experiment. Number of replicates for each experiment is mentioned in figure legends. Data is presented as the mean of replicates and error bars represent standard error of the mean. Number of replicates used in different experiments are also mentioned in Supplementary Figure S3. Test of significance used are one-way ANOVA, two-way ANOVA followed by *post-hoc* Tukey's test ($p < 0.05$), Student's *t*-test and one sample *t*-test. All the statistical analysis was done using SigmaPlot 11.0 (Systat Software, Inc).

RESULTS

Assessment of Bacterial Pathogenicity and Combined Stress Imposition

Chickpea plants infected with Psp and Rs were initially assessed for their pathogenicity by determining *in planta* bacterial number and disease symptom development. Psp inoculated chickpea leaves showed increase in bacterial number till 5 dpi (Figure 1A) and chlorosis was observed on the inoculated leaves at 6 dpi (Figure 1B). These results indicated that Psp is a potential but mild pathogen of chickpea. Rs infiltrated plants showed increase in number of bacterial colony forming units till 5 dpi (Figure 1C), and this was accompanied with appearance of disease symptoms such as yellowing at low bacterial numbers [4.86 Log (cfu/mg)], and wilting and cell death at higher bacterial numbers [6.54 Log (cfu/mg)]. This demonstrated that Rs is also a potential host pathogen of chickpea (Figure 1D).

The effect of combined stress was studied using two methods, viz. (i) drought stress followed by pathogen infection (DP), and (ii) pathogen infection followed by drought stress imposition (PD) as detailed in “methods” section (Supplementary Figures S1–S3, Supplementary Table S2).

Combined Stress Reduced Multiplication of Psp and Decreased Cell Death along with Increased Chlorophyll Content Compared to Infection with Pathogen Alone

As a result of stress interaction, 1.6 and 1.5 fold significant decrease in the bacterial number of Psp was observed under mild and moderate drought stress, respectively, when compared to their number in plants challenged with Psp alone (Figure 2A). The net effect of combined drought and Psp (DPsp) in chickpea was further assessed by determining cell death and total chlorophyll content. In the present study, prominent cell death at 6 dpt was observed in plants individually challenged with both stresses (Figure 2B, Supplementary Figure S4A). A three-fold increase in cell death compared to control sample was observed in leaves of mild drought stress. However, there was no significant change in the extent of cell death observed in response to increase in drought severity. Both moderate and severe drought stressed plants exhibited approximately a two-fold increase in cell death in comparison to control plants (Figure 2B). Psp infection lead to 4.68- and 7-fold increase in cell death compared to mock and absolute control, respectively. Decrease in cell death was observed in DPsp stressed plant when compared to plants infected with Psp alone. However, extent of cell death was similar in both DPsp and drought stressed plants (Figure 2B). The mild DPsp (mild drought with Psp infection), moderate DPsp (moderate drought with Psp infection) and severe DPsp (severe drought with Psp infection) showed 2.2-, 2.05-, and 3.8-fold decrease in cell death, respectively, in comparison to plants challenged with Psp infection alone (Figure 2B).

In case of total chlorophyll content, a significant 1.8 fold decrease was observed in Psp infected leaves than control leaves. In contrary, plants exposed to only drought and DPsp

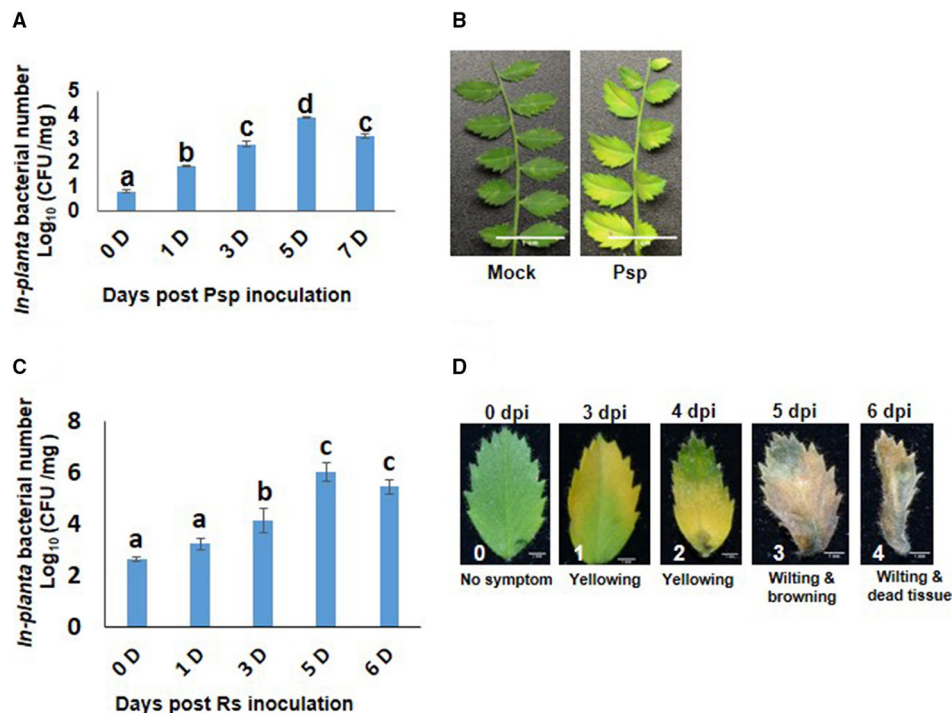


FIGURE 1 | *Pseudomonas syringae* pv. *phaseolicola* (Psp) and *Ralstonia solanacearum* (Rs) multiplication and disease symptoms in chickpea.

Pathogenicity of Psp and Rs in chickpea ICC4958 was assessed by *in planta* bacterial multiplication assay in plants syringe inoculated with 7×10^5 cfu/ml and 2.5×10^6 cfu/ml of bacterial inoculum respectively. The multiplication was studied from 0 to 7 days post inoculation. Data is represented as Log (cfu/mg) and average of 2 replicates and \pm SEM as error bar is plotted for Psp in graph (A) and for Rs in graph (C) is average of 5 replicate. Different letters above bar represents significant difference between means. One-way ANOVA was used for test of significance ($p < 0.05$). Disease symptoms in leaves were observed at 7 days post inoculation in Psp (B). Disease symptoms for Rs was observed from 3 days post infection. The disease symptoms varied from yellowing to cell death to wilting with increase in *in planta* Rs count (D).

stresses did not show any change in their chlorophyll content while compared to control plants (Figure 2C). However, the chlorophyll content was significantly higher in DPsp stressed plants than plants infected with Psp alone. There was 1.7- and 2-fold more chlorophyll in moderate DPsp and severe DPsp stressed plants, respectively, when compared to plants infected with pathogen alone. Chlorophyll content in moderate and severe DPsp was unchanged in comparison to moderate and severe levels of drought stress alone, respectively. However, mild DPsp showed around 2-fold decrease in chlorophyll content over severe drought alone (Figure 2C). The phenotype recorded after 12 dpt showed chlorotic symptoms with disease score of 3.5, 2.3, and 1.6 for Psp alone, mild and moderate DPsp, respectively. Phenotype of severe DPsp was similar to mock control (Supplementary Figures S4B,C). Taken together, the net effect due to combined stress (DP) was similar to drought stress.

Combined Drought and Rs Stress Showed Less Rs Multiplication and Cell Death and More Chlorophyll Content Over Rs Only Stress

The total bacterial count of Rs was constant in chickpea plants exposed to mild drought stress, but their number declined

during severe drought stress as compared to plants infected with Rs only (at 6 dpt; Figure 3A). Similarly, cell death and total chlorophyll content were also assessed in combined stress treated plants as well as their respective controls. Increased cell death was observed in plants individually challenged with drought and Rs infection (Figure 3B, Supplementary Figure S5A). When compared to absolute control, drought stress caused more cell death by 3-, 2.4-, and 2-folds in mild, moderate and severe drought levels, respectively. Rs infection showed 4.15-fold increase in cell death compared to absolute control. During combined stress conditions, increase in cell death was noted in mild DRs (mild drought with Rs) plants compared to absolute control. However, moderate DRs (moderate drought with Rs) and severe DRs (severe drought with Rs) did not have a major impact on the viability of cells as they showed 2- and 1.7-fold reduction in cell death, respectively, compared to Rs alone stress (Figure 3B). Moreover, the extent of cell death observed during moderate and severe DRs, and drought stress alone were similar (Figure 3B).

Total chlorophyll content in the leaves of plants infected with pathogen alone (Rs) was reduced to 2-folds in comparison to control, and it further decreased in mild DR treatment (6-fold reduction compare to control; Figure 3C). There was 1.4-fold decrease in the chlorophyll content of plants challenged with

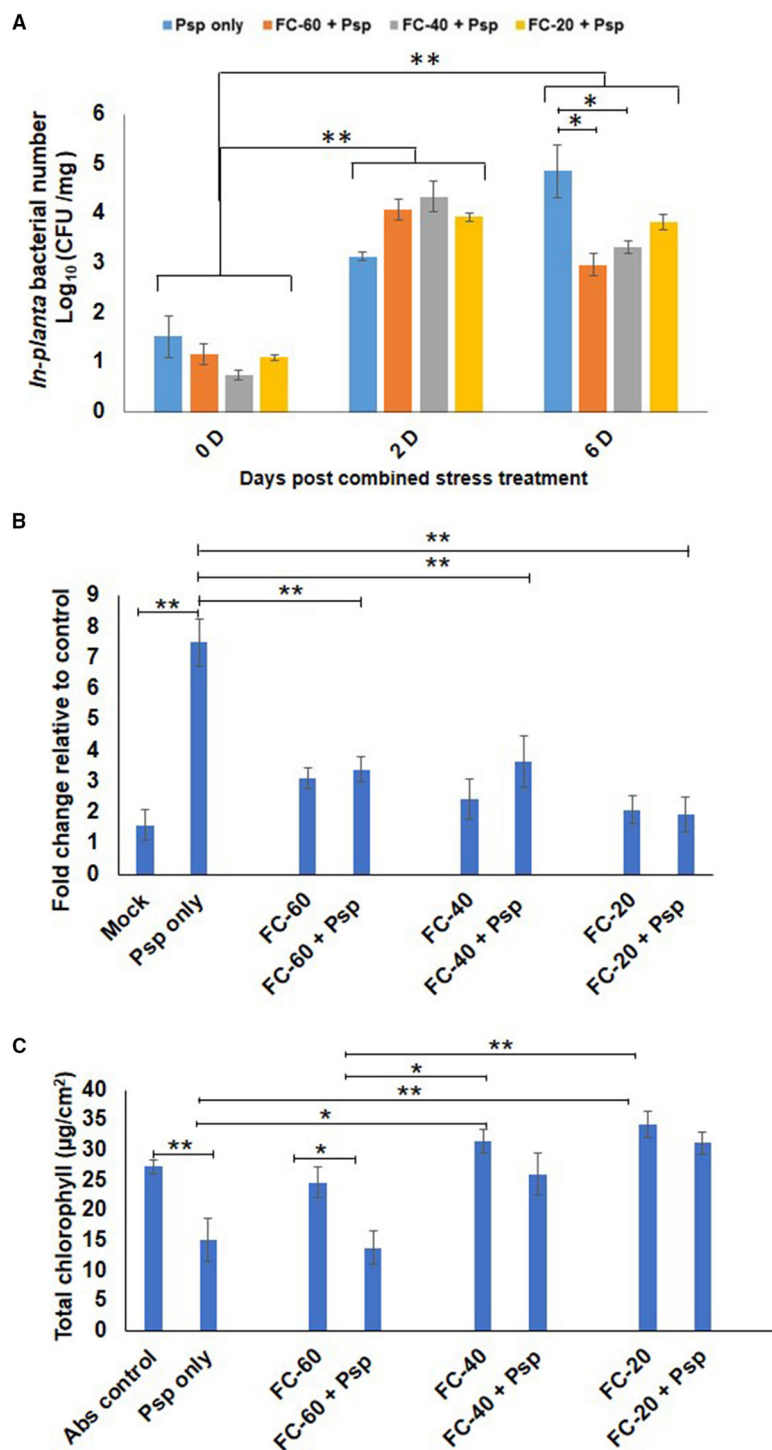


FIGURE 2 | Effect of different levels of drought stress on *P. syringae* pv. *phaseolicola* multiplication and the net effect of combined stress on chickpea.

In-planta number of Psp was measured for the combined stresses plants (mild, moderate, and severe DPsp) and only Psp stress. It was measured for 0, 2, and 6 days post combined stress treatment (dpt). Data is represented as Log (cfu/mg) in graph (A). Each bar is average of 9 replicates and error bar represents \pm SEM. Cell death was studied using trypan blue staining method. The sample was photographed under bright field microscope and intensity of trypan blue was measured using ImageJ software. Graph (B) represents quantitative measurement of cell death in drought and pathogen stressed samples at 6 dpt. Bar represents fold cell death over absolute control and error bar represents \pm SEM, each bar is average of four replicates. Graph (C) represents total chlorophyll content of different combined stress and individual stresses measured at 12 dpt. Each bar represents average of 6 replicates and error bar represents \pm SEM. *, ** represents significant difference at $p < 0.01$ and $p < 0.001$, respectively. Two-way ANOVA was used for test of significance and *post-hoc* Tukey's test was used to represent significant difference between the means.

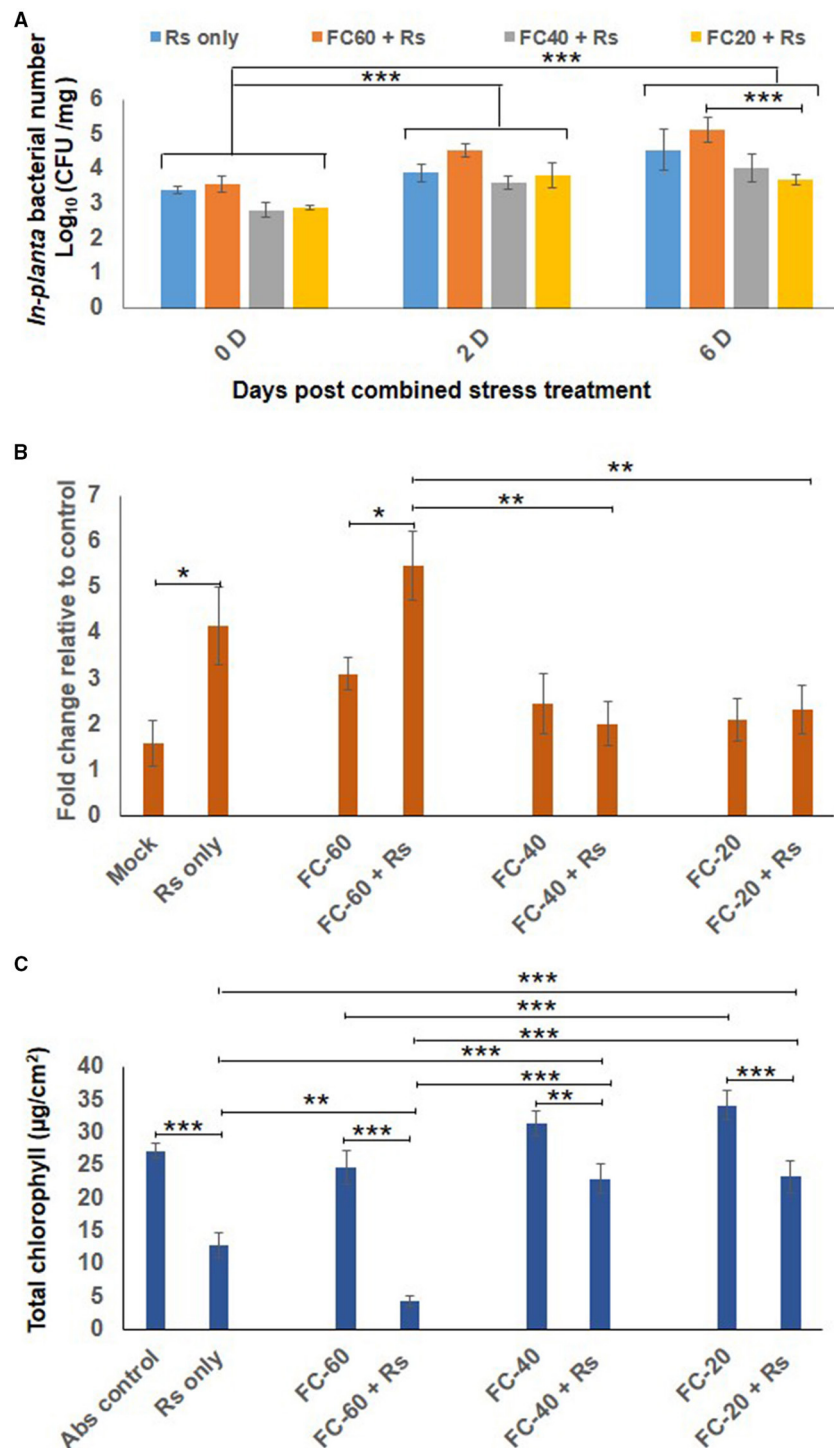


FIGURE 3 | Effect of drought stress on *in planta* multiplication of *R. solanacearum* (Rs) and net effect of combined stress on chickpea. (A) The *in planta* bacterial number of Rs under combined stress (mild, moderate and severe DRs) and only Rs stress condition was measured at 0, 2, and 6 days after combined stress treatment (dpt). Data is represented as Log (cfu/mg) in graph (A). Each bar is average of 9 replicates and error bar represents \pm SEM. Cell death was studied by trypan blue staining method for combined stresses (DRs) as well as individual Rs and drought stresses. The samples were photographed under bright field microscope and intensity of trypan blue was measured using ImageJ software. Graph (B) shows average of four ImageJ intensity value. Y-axis represents fold change over control & error bar signifies \pm SEM. Graph (C) shows total chlorophyll content in leaf disc of chickpea imposed with drought stresses, Rs infection and combined DRs. Chlorophyll estimation was done at 12 dpt. Each bar represents average of six replicates and error bar represents \pm SEM. *, **, and *** represents significant difference at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. Two-way ANOVA was used as test of significance and *post-hoc* Tukey's test was used to calculate significant difference between each mean.

moderate and severe DRs compared to plants at respective drought stress levels. However, the chlorophyll levels in mild and moderate DRs were 1.8-fold high contrasting to plants infected with Rs alone (**Figure 3C**). The difference in chlorophyll content was reflected on the phenotype of plants, as the plants challenged with both mild drought and Rs infection showed increased chlorosis in comparison to plants infected with Rs alone (Supplementary Figure S5C). Chlorotic symptoms with disease scores of 1.6, 4.3, and 1.6 for Rs alone and mild and moderate DRs respectively were recorded at 12 dpt. However, severe DPsp had phenotype similar to mock control with no chlorosis (Supplementary Figures S5B,C). Thus, with mild drought stress, the disease severity was decreased in case of DPsp, but was significantly increased in DRs combined stress.

These results suggest that the net effect of combined stress was more due to the drought stress and also, two pathogens differentially elicited net-effects on plants during combined stress as measured by cell death and chlorophyll content (Supplementary Figures S6A,B). Additionally, our results also indicate that level of drought stress decides elicitation or suppression of plant defenses (Supplementary Figure S6C).

Bacterial Multiplication was Similar in Plants Challenged with PD Combined Stress and Pathogen Stress

Plants infected with pathogen followed by imposition of drought stress (PD) showed similar bacterial number as that of pathogen only treatment (**Figure 4**). During progressive drought in PD stress, plants at 4 days post infection (dpi) experienced 60% FC and at 10 days post infection 20% FC (Supplementary Figure S2B). Therefore, PD stressed plant at 4 and 10 dpi experienced mild and severe combined stress respectively. Combined PD stressed (Psp) and Psp only infected plants showed a constant increase in bacterial count till 10 dpi. In contrary, Rs count increased significantly on 1 dpi in Rs only, but no notable increase was observed on subsequent days (**Figure 4**). However, PspD and RsD did not show significant decrease in bacterial count at mild or severe drought stress. Altogether, the dissimilar effect of DP and PD on bacterial colony number indicates that timing of occurrence of drought stress is important during stress interaction.

Expression of Pathogen Stress Responsive Genes were Differentially Regulated under Combined Stress in Comparison to Individual Stress

Differential expression pattern of drought responsive (*CaNCED1*, 9-cis-epoxycarotenoid dioxygenase; *CaDREB2A*, dehydration responsive element binding; *CaLEA4*, late embryogenesis abundant 4) as well as pathogen responsive (*CaPR4*, thaumatin-like pathogenesis-related protein 4-like; *CaPAL2* phenylalanine ammonia-lyase 2-like) genes were observed during these stress conditions compared to corresponding mock and absolute control at 2 dpt (**Figure 5**). This further validated the stress experienced by the plants. The analysis showed higher expression of drought responsive

genes in all the levels of drought stress. Among these genes, *CaDREB2A* showed 57.5 fold up-regulated expression during drought, whereas the expression of *CaLEA4* and *CaNCED1* increased from 5.6 and 1.7 to 8.4 and 5.4 folds, respectively, with the increase in severity of drought stress from mild to moderate levels (**Figure 5**). The pathogen responsive genes, *CaPR4* and *CaPAL2* showed downregulation during drought stress. Interestingly, both drought and pathogen responsive genes had almost similar expression under mild and severe drought alone stress (**Figure 5**). In case of pathogen challenge, *CaPR4* and *CaPAL2* displayed significant up-regulation in response to both the pathogens; however, *CaPR4* exhibited a relatively higher expression in response to Psp (49.4 fold) than Rs (18.4 fold) (**Figure 5**). Moreover, *CaLEA4* exhibited up-regulated expression pattern in pathogen alone (Psp, Rs) infected plants (**Figure 5**). While in response to DP combined stress, drought responsive genes, *CaDREB2A* and *CaNCED1* showed decreased expression compared to drought alone stress and decreased or similar expression compared to pathogen alone during mild and moderate in both DPsp and DRs stresses. However, *CaLEA4* showed decreased expression in DPsp but increased expression in DRs compared to both the individual stresses. The expression of pathogen stress responsive genes *CaPR4* and *CaPAL2* was downregulated in DPsp compare to Psp alone but they were up-regulated in combined DRs compared to individual stresses (**Figure 5**). Expression of *CaDREB2A*, *CaPR4*, and *CaPAL2* under severe DPsp and DRs was almost similar to respective pathogen alone stress. However, *CaLEA4* and *CaNCED1* showed slightly increased expression compare to pathogen alone.

Expression pattern of stress responsive genes was also studied in the samples where pathogen infection and subsequent drought stress imposition have been performed (PD). Increased expression of *CaLEA2* in response to drought alone, and *CaLEA4* and *CaPAL2* genes in pathogen alone at 10 days post-inoculation of pathogen confirmed the prevalence of drought and pathogen stress in individual stressed plants. Compared to mock control, *CaPAL2* showed lower expression in samples infected with Psp alone, however a four-fold higher expression of this gene was observed in response to infection with Rs alone (**Figure 6**). During combined PspD stress, *CaLEA4*, *CaLEA1*, and *CaLEA2* genes showed higher transcript expression compared to individual drought and pathogen stresses. However, during combined RsD stress, only *CaLEA1* showed 4.5 fold increased expression in comparison to individual stresses. This indicates that the alteration in stress responsive genes was influenced by the nature of infecting pathogen during combined stress response.

DISCUSSION

Simultaneous occurrence of drought and bacterial pathogen infection influences the impact of each other during their interaction *in planta* (Timmusk and Wagner, 1999; McElrone et al., 2003; Mohr and Cahill, 2003; Ramegowda et al., 2013). Moreover, the net impact of combined stress on plants has been reported to be unique compared to individual stresses (Choi et al., 2013). The present study tested the effect of drought stress on the

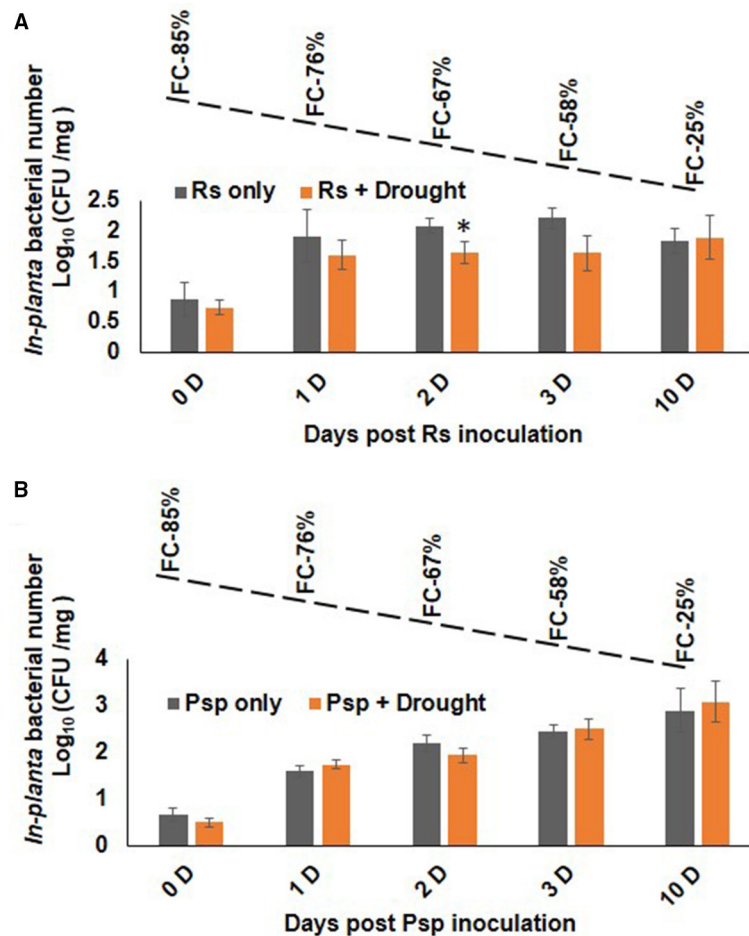


FIGURE 4 | Effect of drought stress on bacterial multiplication under PD combined stress. The *in planta* presence of *R. solanacearum* and *P. syringae* pv. phaseolicola count was measured at 0, 1, 2, 3, and 10 days post infection under RsD (Rs+Drought), only Rs, PspD (Psp+Drought), and only Psp stress. Graph (A) represents the *in planta* count of Rs and graph (B) represents the *in planta* count of Psp. Each bar represents average of 9 biological replicates and error bar represents \pm SEM. *Represents significant difference at $p < 0.05$. Student's *t*-test was used for test of significance. Dotted line denotes gradual decline in soil moisture content with values representing potted plants at actual field capacity (FC) on respective days.

pathogenicity of two different bacterial pathogens in chickpea, and also the net physiological effect by assessing cell death and chlorophyll content during stress combinations (DP and PD).

Our results indicated that pathogen infection preceded by drought stress reduced the multiplication of Psp and Rs in chickpea. There could be three possible reasons for this effect. First, presumably reduced availability of water needed for *in planta* bacterial multiplication. Beattie (2011) has explained that water influences plant-pathogen interaction. Earlier, reduced water potential in bacterial culture media was found to delay the bacterial multiplication (Beattie, 2011). It has also been shown that plants produce localized desiccation at the site of infection and reduced pathogen numbers as a part of basal and effector triggered defenses (Beattie, 2011). In our study, we found reduced leaf RWC in response to drought stress and combined stress (Supplementary Figure S7). Secondly, the drought stress induced molecular and biochemical adaptation in chickpea plants could have contributed to reduced bacterial growth. For example,

drought stress provokes the accumulation of reactive oxygen species (ROS) which at lower concentrations acts as secondary messenger in signal transduction and triggers defense response against pathogen (Lamb and Dixon, 1997). In current study, we observed increased ROS production with increase in drought stress level (Supplementary Figure S8) and therefore priming with drought mediated ROS can be one of the reason for decreased bacterial multiplication. Similarly, PR5 (pathogenesis-related protein-5) and PDF2.1 (plant defensin 1.2) genes, which are known to be involved in pathogen defense (Glazebrook, 2001) are also found to be highly expressed under drought stress (Ramegowda et al., 2013). Boominathan et al. (2004) have shown that drought adaptation increases the expression of serine threonine protein kinases (STPK) which are also involved in pathogen defense (Dangl and Jones, 2001; Zhang et al., 2013). In our study, we found increased expression of *CaPR4* in drought stress alone, and therefore, we assume that it primed the plant for upcoming pathogen stress. Third reason

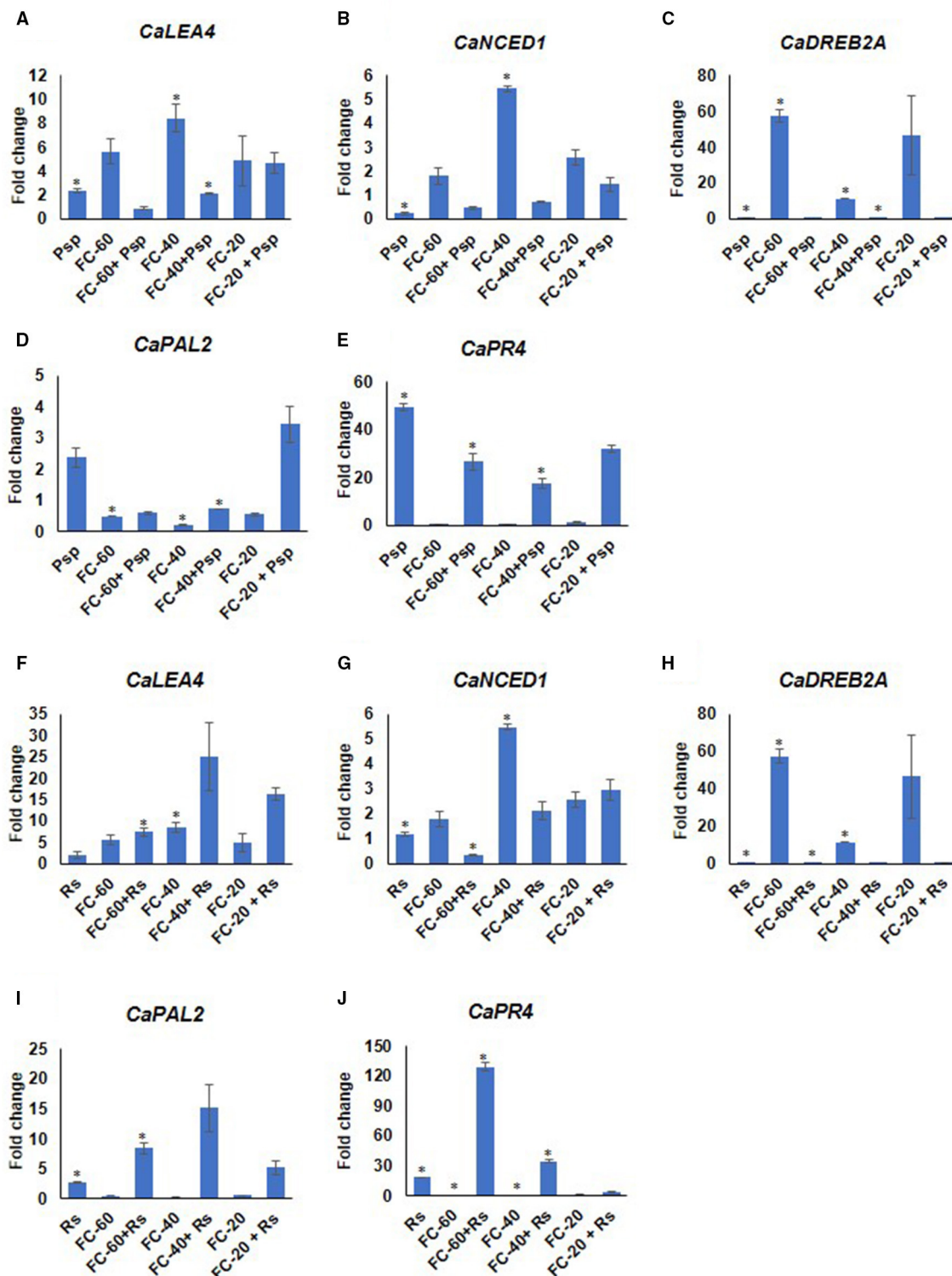


FIGURE 5 | Expression analysis of stress responsive genes of chickpea under drought, pathogen and combined stress (DP). Expression of stress responsive genes in comparison to control was studied using RT-qPCR. The Ct values of different genes were normalized with *Ca18S* internal control. Fold change in gene expression was calculated by $2^{-\Delta\Delta CT}$ method. Mock was considered as reference for DP and pathogen only, and absolute control was reference for drought only. The differential gene expression for *CaLEA4* (A,F) *CaNCED1* (B,G) *CaDREB2A* (C,H) and *CaPAL2* (D,I) and *CaPR4* (E,J) under DPsp and DRs combined stress are represented as bar graph. Each bar signifies average of two biological replicates and error bar represents \pm SEM. Significance was tested by one sample *t*-test. *Denotes significant at $p < 0.05$.

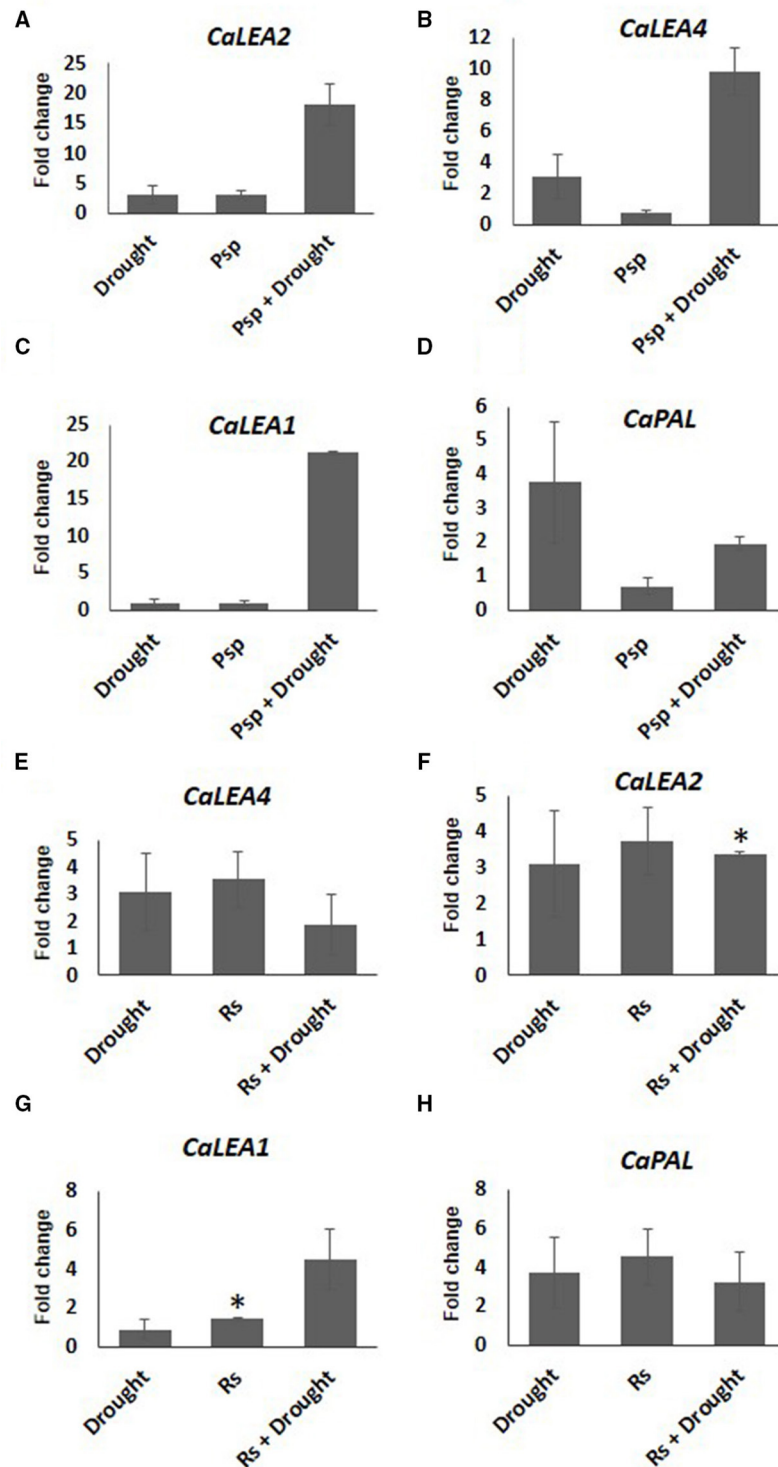


FIGURE 6 | Expression analysis of stress responsive genes of chickpea under drought, pathogen and combined stress (PD). Expression of stress responsive genes in comparison to control was studied using RT-qPCR. The Ct values of different genes were normalized with *CaActin* internal control. Fold change in gene expression was calculated by $2^{-\Delta\Delta CT}$ method. Mock was considered as reference for PD and pathogen only and absolute control was reference for drought only. The differential gene expression for *CaLEA2* (A,F), *CaLEA4* (B,E), *CaLEA1* (C,G) and *CaPAL2* (D,H) under PspD and RsD combined stress are represented as bar graph. Each bar represents average of two biological replicates and error bar represents \pm SEM. Significance was verified by one sample *t*-test. *Denotes significant at $p < 0.05$.

for reduced multiplication could be, physiological, biochemical and molecular mechanism that are unique to combined stress (Pandey et al., 2015). In this study, *CaPAL2* and *CaPR4* showed very high expression under combined stress over individual stresses in DRs. Such fold change was noted to be more than the additive expression of these genes under two individual stresses which could be taken up as unique response by plants under combined stress. Therefore, we assume that this could be one of the reason for the decreased multiplication under combined stress.

During DRs, multiplication of Rs was found to be decreased at severe drought level. However, mild drought stress did not reduce the Rs multiplication. This indicates that intensity of drought stress plays important role in combined stress effect. During stress interaction, the net outcome of the stress response decides whether the plant is capable of circumventing the combined stress effect or not. In the present study, we found that the drought stress (individual) leads to increased cell death but does not affect chlorophyll content of the plant, whereas pathogen infection lead to increase in cell death and disease associated decrease in chlorophyll content. However, the impact of DPsp stress on cell death and chlorophyll was similar to drought stress and it was reduced in comparison to only Psp stress. Similarly, the net effect of DRs except mild DRs on cell death was almost similar to drought stress and reduced in comparison to only Rs stress. This indicated that drought stressed plants were able to defend themselves better against upcoming pathogen.

In conclusion, the study demonstrates that priming of drought stress reduces the multiplication of Psp and Rs pathogens

in planta. The net effect of combined stress was not additive and drought has more impact during combined occurrence with pathogen. The study also shows that the outcome of combined stress is conditional and depends on which stress factor occurs first in the plant.

AUTHOR CONTRIBUTIONS

MS conceived the idea and provided resources. MS, RS designed the experiments. RS, AG performed the experiments. RS analyzed the data under the guidance of MS. MS, RS wrote the paper. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

Projects at MS lab are supported by DBT-Ramalingaswami re-entry fellowship grant (BT/RLF/re-entry/23/2012). RS thank Science and Engineering Research Board (SERB), Department of Science and Technology (DST) for providing fellowship and research grant (SB/YS/LS-237/2013). Authors thank Mr. Mehanathan Muthamilarasan for critical reading of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00902>

REFERENCES

- Arnon, D. (1949). Copper enzymes in isolated chloroplasts, polyphenoxidase in beta vulgaris. *Plant Physiol.* 24, 1–15. doi: 10.1104/pp.24.1.1
- Beattie, G. A. (2011). Water relations in the interaction of foliar bacterial pathogens with plants. *Annu. Rev. Phytopathol.* 49, 533–555. doi: 10.1146/annurev-phyto-073009-114436
- Boominathan, P., Shukla, R., Kumar, A., Manna, D., Negi, D., Verma, P. K., et al. (2004). Long term transcript accumulation during the development of dehydration adaptation in *Cicer arietinum*. *Plant Physiol.* 135, 1608–1620. doi: 10.1104/pp.104.043141
- Choi, H.-K., Iandolino, A., Goes da Silva, F., and Cook, D. R. (2013). Water deficit modulates the response of *Vitis vinifera* to the Pierce's disease pathogen *Xylella fastidiosa*. *Mol. Plant Microbe Interact.* 26, 643–657. doi: 10.1094/MPMI-09-12-0217-R
- Dangl, J. L., and Jones, J. D. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- Desprez-Loustau, M. L., Robin, C., Reynaud, G., Déqué, M., Badeau, V., Piou, D., et al. (2007). Simulating the effects of a climate-change scenario on the geographical range and activity of forest-pathogenic fungi. *Can. J. Plant. Pathol.* 29, 101–120. doi: 10.1080/07060660709507447
- Gaur, P. M., Jukanti, A. K., and Varshney, R. K. (2012). Impact of genomic technologies on chickpea breeding strategies. *Agronomy* 2, 199–221. doi: 10.3390/agronomy2030199
- Genin, S. (2010). Molecular traits controlling host range and adaptation to plants in *Ralstonia solanacearum*. *New Phytol.* 187, 920–928. doi: 10.1111/j.1469-8137.2010.03397.x
- Glazebrook, J. (2001). Genes controlling expression of defense responses in *Arabidopsis* — 2001 status. *Curr. Opin. Plant Biol.* 4, 301–308. doi: 10.1016/S1369-5266(00)00177-1
- Gupta, A., Dixit, S. K., and Senthil-Kumar, M. (2016a). Drought stress predominantly endures *Arabidopsis thaliana* to *Pseudomonas syringae* infection. *Front. Plant Sci.* 7:808. doi: 10.3389/fpls.2016.00808
- Gupta, A., Sarkar, A. K., and Senthil-Kumar, M. (2016b). Global transcriptional analysis reveals unique and shared responses in *Arabidopsis thaliana* exposed to combined drought and pathogen stress. *Front. Plant Sci.* 7:686. doi: 10.3389/fpls.2016.00686
- Hiscox, J. D., and Israelstam, G. F. (1980). Erratum: a method for the extraction of chlorophyll from leaf tissue without maceration. *Can. J. Bot.* 58, 403–403. doi: 10.1139/b80-044
- Jha, U. C., Chaturvedi, S. K., Bohra, A., Basu, P. S., Khan, M. S., and Barh, D. (2014). Abiotic stresses, constraints and improvement strategies in chickpea. *Plant Breed* 133, 163–178. doi: 10.1111/pbr.12150
- Kanda, A., Yasukochi, M., Ohnishi, K., Kiba, A., Okuno, T., and Hikichi, Y. (2003). Ectopic expression of *Ralstonia solanacearum* effector protein PopA early in invasion results in loss of virulence. *Mol. Plant Microbe Interact.* 16, 447–455. doi: 10.1094/MPMI.2003.16.5.447
- Kiba, A., Tomiyama, H., Takahashi, H., Hamada, H., Ohnishi, K., Okuno, T., et al. (2003). Induction of resistance and expression of defense-related genes in tobacco leaves infiltrated with *Ralstonia solanacearum*. *Plant Cell Physiol.* 44, 287–295. doi: 10.1093/pcp/pcg037
- Koch, E., and Slusarenko, A. (1990). *Arabidopsis* is susceptible to infection by a downy mildew fungus. *Plant Cell* 2, 437–445. doi: 10.1105/tpc.2.5.437
- Lamb, C., and Dixon, R. A. (1997). The oxidative burst in plant disease resistance. *Annu. Mol. Biol.* 48, 251–275. doi: 10.1146/annurev.arplant.48.1.251
- Li, H., Rodda, M., Gnanasambandam, A., Aftab, M., Redden, R., Hobson, K., et al. (2015). Breeding for biotic stress resistance in chickpea: progress and prospects. *Euphytica* 204, 257–288. doi: 10.1007/s10681-015-1462-8
- Liu, X., Sun, Y., Körner, C. J., Du, X., Vollmer, M. E., and Pajeroska-Mukhtar, K. M. (2015). Bacterial leaf infiltration assay for fine characterization of

- plant defense responses using the *Arabidopsis thaliana*-*Pseudomonas syringae* Pathosystem. *J. Vis. Exp.* 104:e53364. doi: 10.3791/53364
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Mattson, W. J., and Haack, R. A. (1987). The role of drought in outbreaks of plant-eating insects. *Bioscience* 37, 110–118. doi: 10.2307/1310365
- McElrone, A. J., Sherald, J. L., and Forseth, I. N. (2003). Interactive effects of water stress and xylem-limited bacterial infection on the water relations of a host vine. *J. Exp. Bot.* 54, 419–430. doi: 10.1093/jxb/erg046
- Mohr, P. G., and Cahill, D. M. (2003). Absciscic acid influences the susceptibility of *Arabidopsis thaliana* to *Pseudomonas syringae* pv. tomato and *Peronospora parasitica*. *Funct. Plant Biol.* 30, 461–469. doi: 10.1071/FP02231
- Nene, Y. L., Reddy, M. V., Haware, M. P., Ghanekar, A. M., Amin, K. S., Pande, S., et al. (2012). *Field Diagnosis of Chickpea Diseases and their Control. Information Bulletin No. 28 (revised)*. Technical Report, International Crops Research Institute for the Semi-Arid Tropics, Patancheru.
- Pandey, P., Ramegowda, V., and Senthil-Kumar, M. (2015). Shared and unique responses of plants to multiple individual stresses and stress combinations: physiological and molecular mechanisms. *Front. Plant Sci.* 6:723. doi: 10.3389/fpls.2015.00723
- Prasch, C. M., and Sonnewald, U. (2013). Simultaneous application of heat, drought, and virus to *Arabidopsis* plants reveals significant shifts in signaling networks. *Plant Physiol.* 162, 1849–1866. doi: 10.1104/pp.113.221044
- Ramegowda, V., and Senthil-Kumar, M. (2014). The interactive effects of simultaneous biotic and abiotic stresses on plants: Mechanistic understanding from drought and pathogen combination. *J. Plant Physiol.* 176, 47–54. doi: 10.1016/j.jplph.2014.11.008
- Ramegowda, V., Senthil-Kumar, M., Ishiga, Y., Kaundal, A., Udayakumar, M., and Mysore, K. S. (2013). Drought stress acclimation imparts tolerance to *Sclerotinia sclerotiorum* and *Pseudomonas syringae* in *Nicotiana benthamiana*. *Int. J. Mol. Sci.* 14, 9497–9513. doi: 10.3390/ijms14059497
- Reynolds, S. G. (1970). The gravimetric method of soil moisture determination Part I A study of equipment, and methodological problems. *J. Hydrol.* 11, 258–273. doi: 10.1016/0022-1694(70)90066-1
- Saettler, A. W. (1991). *Angular Leaf Spot. Compendium of Bean Diseases* St. Paul, MN: American Phytopathological Society Press.
- Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. doi: 10.1038/nmeth.2089
- Sharma, M., and Pande, S. (2013). Unravelling effects of temperature and soil moisture stress response on development of dry root rot [*Rhizoctonia bataticola* (Taub.)] Butler in Chickpea. *Am. J. Plant Sci.* 04, 584–589. doi: 10.4236/ajps.2013.43076
- Shinohara, R., Kanda, A., Ohnishi, K., Kiba, A., and Hikichi, Y. (2005). Contribution of folate biosynthesis to *Ralstonia solanacearum* proliferation in intercellular spaces. *Appl. Environ. Microbiol.* 71, 417–422. doi: 10.1128/AEM.71.1.417-422.2005
- Suzuki, N., Rivero, R. M., Shulaev, V., Blumwald, E., and Mittler, R. (2014). Tansley review Abiotic and biotic stress combinations. *New Phytol.* 203, 32–43. doi: 10.1111/nph.12797
- Timmusk, S., and Wagner, E. G. (1999). The plant-growth-promoting rhizobacterium *Paenibacillus polymyxa* induces changes in *Arabidopsis thaliana* gene expression: a possible connection between biotic and abiotic stress responses. *Mol. Plant Microbe Interact.* 12, 951–959. doi: 10.1094/MPMI.1999.12.11.951
- Vaillau, F., Sartorel, E., Jardinaud, M.-F., Chardon, F., Genin, S., Huguet, T., et al. (2007). Characterization of the interaction between the bacterial wilt pathogen *Ralstonia solanacearum* and the model legume plant *Medicago truncatula*. *Mol. Plant Microbe Interact.* 20, 159–167. doi: 10.1094/MPMI-20-2-0159
- Wang, K., Senthil-Kumar, M., Ryu, C.-M., Kang, L., and Mysore, K. S. (2012). Phytosterols play a key role in plant innate immunity against bacterial pathogens by regulating nutrient efflux into the apoplast. *Plant Physiol.* 158, 1789–1802. doi: 10.1104/pp.111.189217
- Yadeta, K. A., and Thomma, B. P. H. J. (2013). The xylem as battleground for plant hosts and vascular wilt pathogens. *Front. Plant Sci.* 4:97. doi: 10.3389/fpls.2013.00097
- Yáñez-López, R., Torres-Pacheco, I., Guevara-González, R. G., Hernández-Zul, M. I., Quijano-Carranza, J., and Rico-García, E. (2012). The effect of climate change on plant diseases. *African J. Biotechnol.* 11, 2417–2428. doi: 10.5897/AJB10.2442
- Zhang, Y., Wang, X., Li, Y., Wu, L., Zhou, H., Zhang, G., et al. (2013). Ectopic expression of a novel Ser/Thr protein kinase from cotton (*Gossypium barbadense*), enhances resistance to *Verticillium dahliae* infection and oxidative stress in *Arabidopsis*. *Plant Cell Rep.* 32, 1703–1713. doi: 10.1007/s00299-013-1481-7
- Zhu, H., Choi, H.-K., Cook, D. R., and Shoemaker, R. C. (2005). Bridging model and crop legumes through comparative genomics. *Plant Physiol.* 137, 1189–1196. doi: 10.1104/pp.104.058891

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Sinha, Gupta and Senthil-Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Shoot and Root Traits Contribute to Drought Resistance in Recombinant Inbred Lines of MD 23–24 × SEA 5 of Common Bean

Jose Polania *, Idupulapati M. Rao *, Cesar Cajiao, Miguel Grajales, Mariela Rivera, Federico Velasquez, Bodo Raatz and Stephen E. Beebe

Centro Internacional de Agricultura Tropical, Cali, Colombia

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Consejo Superior de Investigaciones
Científicas, Spain

Reviewed by:

Carla Pinheiro,
Faculdade de Ciências e Tecnologia
da Universidade Nova de Lisboa,
Portugal
Hussein Shimelis,
University of KwaZulu-Natal,
South Africa

*Correspondence:

Jose Polania
j.a.polania@cgiar.org
Idupulapati M. Rao
i.rao@cgiar.org

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 02 December 2016

Accepted: 17 February 2017

Published: 03 March 2017

Citation:

Polania J, Rao IM, Cajiao C,
Grajales M, Rivera M, Velasquez F,
Raatz B and Beebe SE (2017) Shoot
and Root Traits Contribute to Drought
Resistance in Recombinant Inbred
Lines of MD 23–24 × SEA 5 of
Common Bean.
Front. Plant Sci. 8:296.
doi: 10.3389/fpls.2017.00296

Drought is the major abiotic stress factor limiting yield of common bean (*Phaseolus vulgaris* L.) in smallholder systems in Latin America and eastern and southern Africa; where it is a main source of protein in the daily diet. Identification of shoot and root traits associated with drought resistance contributes to improving the process of designing bean genotypes adapted to drought. Field and greenhouse studies were conducted at the International Center for Tropical Agriculture (CIAT), Palmira, Colombia to determine the relationship between grain yield and different shoot and root traits using a recombinant inbred lines (RILs) population (MD23–24 × SEA 5) of common bean. The main objectives of this study were to identify: (i) specific shoot and root morpho-physiological traits that contribute to improved resistance to drought and that could be useful as selection criteria in breeding beans for drought resistance; and (ii) superior genotypes with desirable shoot and root traits that could serve as parents in breeding programs that are aimed at improving drought resistance. A set of 121 bean genotypes (111 RILs, 2 parents, 8 checks) belonging to the Mesoamerican gene pool and one cowpea variety were evaluated under field conditions with two levels of water supply (irrigated and rainfed) over three seasons. To complement field studies, a greenhouse study was conducted using plastic cylinders with soil inserted into PVC pipes, to determine the relationship between grain yield obtained under field conditions with different root traits measured under greenhouse conditions. Resistance to drought stress was positively associated with a deeper and vigorous root system, better shoot growth, and superior mobilization of photosynthates to pod and seed production. The drought resistant lines differed in their root characteristics, some of them with a vigorous and deeper root system while others with a moderate to shallow root system. Among the shoot traits measured, pod harvest index, and seed number per area could serve as useful selection criteria for assessing sink strength and for genetic improvement of drought resistance in common bean.

Keywords: deep rooting, intermittent drought, pod harvest index, root system, seed number

INTRODUCTION

Drought is the main abiotic constraint of common bean (*Phaseolus vulgaris* L.) affecting around 60% of bean producing regions and causing 10–100% reduction in production (Rao, 2014; Polania et al., 2016c). Beans are cultivated by small farmers in Latin America and eastern and southern Africa, under unfavorable climate conditions and minimum use of inputs (Beebe et al., 2014; Rao, 2014). It is expected that the world demand for legumes will increase in the future, not only in developing countries, but also in the developed nations given the trend toward healthier diets (Daryanto et al., 2015). Common beans have to confront climate change and the associated increase of temperature and evapotranspiration together with erratic and lower rainfall (Beebe et al., 2013; Polania et al., 2016a,b; Rippke et al., 2016). The development of bean varieties resistant to drought stress conditions through breeding to ensure food security in marginal areas is a useful strategy to face these new challenges.

Conventional breeding for improving resistance to drought has been based essentially in the selection of the superior genotypes in grain yield (GY) under drought stress (Rosales et al., 2012); with low consideration for defining the physiological basis of drought resistance. The integration and the understanding of the physiological basis of yield limitations due to drought stress, will contribute to developing physiological selection tools to support plant breeding programs (Araus et al., 2002; Girdthai et al., 2009; Mir et al., 2012). Some of the benefits from improved understanding and use of the physiological traits and mechanisms would be the possibility of combining parents with complementary traits, resulting in additive gene action for improving drought resistance (Reynolds and Trethowan, 2007; Mir et al., 2012).

Phenotypic characterization for drought resistance has resulted in the identification of some morpho-physiological traits and process related to improved drought resistance. Processes that are known to influence drought resistance include: more acquisition of water by the root system from the soil profile to facilitate transpiration, greater production of canopy biomass (CB) and an efficient and increased mobilization of accumulated carbon to the harvestable product (Passioura, 1997; Condon et al., 2004; Polania et al., 2016a; Rao et al., 2016a). Several traits have been reported to improve resistance to drought, and their contribution to superior GY depends on the type of drought (early, intermittent, and terminal) and the agro-ecological conditions where the crop is planted. Ideotypes and plant models have been developed for targeting in plant breeding according to agro-ecological zones and types of drought; for example, the isohydric (“water saving”) plant model and the anisohydric (“water spending”) plant model. The water saving ideotype might have an advantage in the harsh environments, whereas the water spending ideotype will perform relatively better under more moderate drought conditions (Blum, 2015; Polania et al., 2016a).

It has been reported that traits related with higher water use efficiency (WUE) and conserving water at vegetative stage (lower leaf conductance, smaller leaf canopy), would make more water

available for reproductive growth and grain filling, resulting in better grain yield under terminal drought stress conditions (Zaman-Allah et al., 2011; Araújo et al., 2015). Increased WUE, could have a penalty in GY, to reduce the rate of transpiration and crop water use, processes that are crucial for carbon assimilation (Blum, 2009; Sinclair, 2012). Blum (2009) proposed the term, effective use of water (EUW), which implies maximal soil moisture capture for transpiration, and also involves decreased non-stomatal transpiration and minimal water loss by soil evaporation. In the water spending model, the EUW would be the main component to consider in plant breeding program for drought resistance, and it is relevant when there is still soil water available at maturity or when deep-rooted genotypes access water deep in the soil profile that is not normally available (Araus et al., 2002; Polania et al., 2016a).

Previous research on common bean under drought stress has suggested the relevance of “water spending” model for improving drought resistance through EUW. Positive relationships have been observed between GY and carbon isotope discrimination (CID) and also with root length density in different genotypes grown under drought stress; indicating that plants under drought stress access more water, resulting in increased stomatal conductance and higher GY (Sponchiado et al., 1989; White et al., 1990; White, 1993; Hall, 2004; Polania et al., 2012, 2016a). Phenotypic evaluation of root traits under drought stress has shown the contribution of deep rooting to access more water from deeper soil layers (Sponchiado et al., 1989; White and Castillo, 1992; Polania et al., 2009, 2012; Beebe et al., 2013, 2014; Lynch, 2013; Rao, 2014; Rao et al., 2016b) and also increased production of fine roots in top soil to take advantage of intermittent rains (Eissenstat, 1992; Huang and Fry, 1998; Polania et al., 2009; Butare et al., 2011; Lynch, 2013; Beebe et al., 2014; Rao et al., 2016a,b).

Increased water extraction capacity and higher crop growth must be accompanied by an improved harvest index (HI) to increase drought resistance. Better remobilization of photosynthates to grain production is needed for the success of superior genotypes under stress. In the case of common bean, the contribution of superior remobilization of reserves from vegetative plant structures to pod and seed formation have been widely documented (Assefa et al., 2013; Beebe et al., 2013; Rao et al., 2013, 2016a; Rao, 2014; Polania et al., 2016a,c). Two dry matter partitioning indices have been shown to be relevant to improved drought resistance: pod partitioning index (PPI) which indicates the extent of mobilization of assimilates from the vegetative structures to pod formation, and pod harvest index (PHI) which indicates the extent of mobilization of assimilates from the pod wall to grain formation (Rao et al., 2013). Photosynthate supply could exert significant and positive quantitative influence on sink strength through setting of grain and pod numbers because abortion rates were shown to be positively related to seed size (Lord and Westoby, 2012).

The strategic combination of specific shoot and root traits seems to be the key in improved resistance to drought in common beans, and no single trait was identified for its unique and dominant contribution to drought resistance (Polania et al., 2016a,c). For that reason, it is relevant to evaluate different shoot

and root traits in a same group of genotypes under drought stress as well as under optimal conditions. This will allow to identify the traits that are not only contributing to improved drought resistance but also responding to irrigation. Most of the lines identified with superior drought resistance in common bean are from the Mesoamerican gene pool, where some lines from Durango race were found to be far superior in their drought resistance (Beebe et al., 2013). For example, the drought resistant line SEA 5 is derived from Durango race and it showed greater ability for remobilization of photoassimilates contributing to higher GY under drought conditions (Beebe et al., 2013; Polania et al., 2016c; Rao et al., 2016a).

The main objectives of this study were to identify: (i) specific shoot and root morpho-physiological traits that contribute to improved resistance to drought and that could be useful as selection criteria in breeding beans for drought resistance; and (ii) superior genotypes with desirable shoot and root traits that could serve as parents in breeding programs that are aimed to improve drought resistance.

MATERIALS AND METHODS

Plant Material

For this study a total of 121 (but 118 for comparison) bean genotypes belonging to the Mesoamerican gene pool and one cowpea variety were selected: 111 RILs of MD 23–24 × SEA 5, two parents (MD 23–24 and SEA 5), and eight checks [Cowpea cv. Mouride, Tio Canela 75, DOR 390, EAP 9510-77, SEA 5 (twice), MD 23–24, and SEA 15]. Cowpea genotype was included for relative comparison of common bean with cowpea for drought resistance. The line MD 23–24 is superior in commercial grain quality and it is also known as “Bribri,” it is small red bean, developed by the Escuela Agrícola Panamericana (EAP), Zamorano, Honduras, and released as a good yielding, well adapted to low soil fertility, and disease resistant cultivar (Rosas et al., 2003). The CIAT bred line SEA 5 is very well adapted to drought, it has small (22–25 g 100 seed⁻¹) cream-colored seeds and Type III growth habit; also resistant to *Fusarium* root rot and has the I gene for resistance to bean common mosaic virus (BCMV). It is susceptible to anthracnose, common bacterial blight, and rust (Singh et al., 2001). The progenies from the cross were advanced by bulk method up to F4 generation and then to two more generations (F5, F6) by pedigree method followed by bulking in F7 generation.

Shoot Phenotyping under Field Conditions

Experimental Site and Meteorological Conditions

Three field trials were conducted during the dry season (from June to September in each year of 2003, 2004 and 2007), at the main experiment station of the International Center for Tropical Agriculture (CIAT) in Palmira, Colombia, located at 3° 29' N latitude, 76° 21' W longitude and an altitude of 965 masl. Basic characteristics of this field site were described previously (Beebe et al., 2008). The soil is a Mollisol (Aquic Hapludoll) with adequate nutrient supply and is estimated to permit storage of 100 mm of available water (assuming 1.0 m of effective root growth with −0.03 and −1.5 MPa upper and lower limits for

soil matric potential). During the crop-growing season in field conditions, maximum and minimum air temperatures in 2003 were 33 and 14.6°C, in 2004 were, 34.4 and 15.6°C and in 2007 were, 30.5 and 18.6°C, respectively (Figure 1). The total rainfall during the active crop growth was 126.5 mm in 2003, 110.4 mm in 2004 and 243.1 mm (a significant proportion of which fell during seed filling) in 2007. The pan evaporation was of 363 mm in 2003, 390 mm in 2004, and 431 mm in 2007. These data on rainfall and pan evaporation together with rainfall distribution indicated that the crop suffered intermittent drought in all 3 years during active growth and development. Two levels of water supply (irrigated and rainfed) were applied to simulate well-watered (control) and drought stress treatments. Trials were furrow irrigated up to 100% field capacity (approximately 35 mm

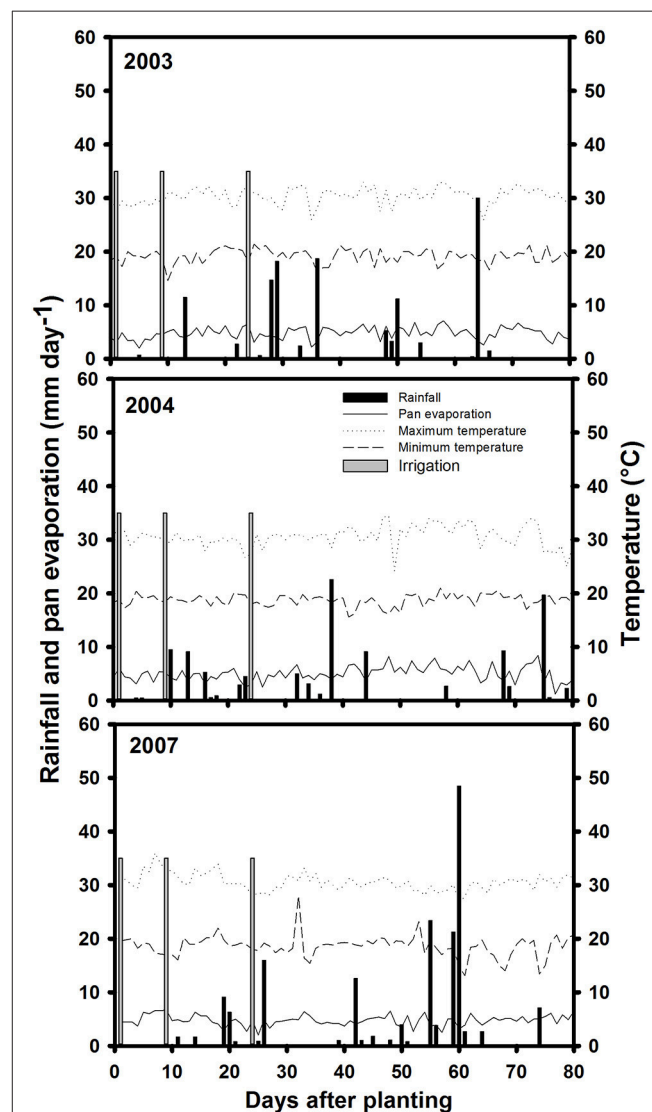


FIGURE 1 | Rainfall distribution and irrigation application, pan evaporation, maximum, and minimum temperatures during crop growing period at Palmira during 2003, 2004, and 2007 crop growing seasons.

of water per irrigation). The drought stress treatment under rainfed conditions received irrigations at 3 days before planting and at 9 and 23 days after planting. Irrigation was suspended after the third irrigation to induce drought stress conditions. For the non-stress or irrigated (control) treatment, the crop was irrigated until physiological maturity with a total of six irrigations in 2003 and seven irrigations in both 2004 and 2007.

Experimental Design

We used an 11×11 partially balanced lattice design with three replications in all three seasons. Details on planting and management of the trial were similar to those reported before (Beebe et al., 2008). Experimental units consisted of 4 rows, 3.72 m long by 0.6 m wide with 7 cm between plants in the row (equivalent to 24 plants m^{-2}). Trials were managed by controlling weeds with application of herbicides (Fomesafen, Fluazifop-p-butyl, and Bentazon) and pests and diseases by spraying with insecticides (Thiametoxam, Clorpirifos, Imidacloprid, Abamectina, Cyromazine, and Milbemectin) and fungicides (Benomil and Carboxin) as needed.

Yield Measurements and Phenological Assessment

Grain was harvested from two central rows after discarding end plants in both the irrigated and drought plots. In order to compare shoot dry biomass with grain dry weight and to quantify dry matter distribution among plant parts, mean values of grain yield per hectare were corrected for 0% moisture in grain. Days to physiological maturity (DPM) were determined for each plot as the number of days after planting until 50% of plants have at least one pod losing its green pigmentation.

Physiological Measurements under Field Conditions

At mid-pod filling, a 50 cm segment of the row (equivalent to an area of 0.3 m^2) from each plot with about 7 plants was used for destructive sampling to measure leaf area index (LAI), canopy biomass (CB) and dry matter distribution between leaves, stems and pods. Leaf area was measured using a leaf area meter (model LI-3000, LI-COR, NE, USA) and the leaf area index (LAI) was calculated. Also, at mid-pod filling SPAD chlorophyll meter readings (SCMR) were made on a fully expanded young leaf of three different plants within each replication by using a non-destructive, hand-held chlorophyll meter (SPAD-502 Chlorophyll Meter). At the time of harvest, plants in 50 cm of a row from each plot were cut and dry weights of stem, pod, seed, pod wall, seed number per area (SNA) per m^2 , and pod number per area (PNA) per m^2 were recorded. The following attributes were determined according to Beebe et al. (2013): harvest index (HI) (%): seed biomass dry weight at harvest/total shoot biomass dry weight at mid-pod filling $\times 100$; pod harvest index (PHI) (%): seed biomass dry weight at harvest/pod and seed biomass dry weight at harvest $\times 100$; PPI (%): pod and seed biomass dry weight at harvest/total shoot biomass dry weight at mid-pod filling $\times 100$. HI and PPI were estimated using the CB-value at mid-pod filling growth stage which is assumed to be the time that reflects the maximum vigor of the genotype; from this time common bean begins to lose CB through leaf fall, especially under drought stress.

Root Phenotyping Using Soil Cylinder System

Experimental Conditions

A greenhouse study was conducted at CIAT using an Andisol from the region of Darien, Colombia, and mixed (2:1 w/w) with river sand. Soil cylinders were carefully packed with soil:sand mixture, with a final bulk density of 1.2 g cm^{-3} . Soil was fertilized with adequate levels of nutrients (kg/ha: 40 N, 50 P, 100 K, 101 Ca, 29 Mg, 20 S, 2 Zn, 2 Cu, 0.1 B, and 0.1 Mo) at planting by mixing with the soil. The seeds were germinated and uniform seedlings were selected for transplanting to transparent plastic cylinders (80 cm long, 7.5 cm diameter), each of which was inserted into PVC sleeve-tubes (Polania et al., 2009). Plants were grown for 48 days in these soil cylinders with an average maximum and minimum temperature of 34° and 21°C .

Experimental Design

A randomized complete block design (RCB) with three replications was used. Two water supply treatments were applied: (1) well-watered (WW) at 80% field capacity and (2) progressive water stress (WS) with no watering after 10 days of growth in order to simulate terminal drought stress conditions. The initial soil moisture for all the treatments was at 80% of field capacity. The plants with well-watered treatment were maintained close to 80% field capacity by weighing each cylinder every 2 days and applying water to the soil at the top of the cylinder. Plants with progressive soil drying treatment received no water application and each cylinder was weighed at 2 day intervals for the determination of decrease in soil moisture content until the time of plant harvest.

Physiological Measurements under Greenhouse Conditions

Plants were harvested at 48 days after transplant (38 days of withholding of water application in the case of drought treatment). Visual rooting depth was measured during the experiment at 7 day intervals using a ruler with cm scale, registering the total length reached by the visible roots of the plastic cylinder. At harvest, leaf area (LICOR model LI-3000), shoot biomass distribution and root distribution were measured. For root distribution traits, the cylinder was sliced into six layers (0–5, 5–10, 10–20, 20–40, 40–60, and 60–75 cm) and the roots in each soil layer were washed free of soil and sand. The washed roots were scanned as images by a desk scanner. From the scanned images, total root length ($m \text{ plant}^{-1}$), and fine roots proportion (%), were measured using image analysis by WinRHIZO software (Regent Instruments Inc., Quebec, Canada). Root and shoot dry weight was determined after the root and shoot samples were dried in an oven at 60°C for 48 h.

Statistical Analysis

Statistical analysis was performed using the Mixed Procedure of SAS V9.2 (SAS Institute Inc, 2008) and Sigma plot. For the purposes of estimating adjusted line means and comparing check entries with experimental lines, entries, environments, and their interactions were considered fixed effects. Replications and blocks within replications were considered random effects.

The mixed model uses information on means of fixed effects that are contained in the differences between blocks, combining the traditional information within the block (Proc GLM) with the new information between blocks (Proc Mixed). Significance of differences among genotypes was tested by the Tukey-Kramer method. Although the mixed model does not present an overall standard error, given that the standard errors associated with each genotype were very similar for each variable, an average standard error was calculated to estimate a significant difference for each variable in the two treatments, for the purpose of visualizing comparisons. The relationships between selected parameters were investigated using the Pearson's correlation test (level of probability at 0.05, 0.01, and 0.001). The principal component analysis (PCA) was used to determine the relationship between multiple variables using PRINCOMP of SAS V9.2 (SAS Institute Inc, 2008). PCA permits creating values that reflect the combined effect of multiple variables that are acting in a similar way.

RESULTS

Phenotypic Evaluation of Shoot Traits under Field Conditions

The data on rainfall distribution, irrigation application, and pan evaporation in both trials indicated that the crop suffered intermittent drought stress during crop development under rainfed conditions (Figure 1). The mean value of GY under drought stress conditions was 1,181 kg ha⁻¹ compared with the mean irrigated GY of 1,845 kg ha⁻¹ with about 36% reduction of mean grain yield under drought stress (Figure 2). Under drought stress conditions in the field, the GY of 118 genotypes ranged from 690 to 1,575 kg ha⁻¹ (Figure 2). Among the lines tested, three RILs, MR 81, MR 112, and MR 25 were outstanding in their adaptation to drought stress conditions. These three lines were also responsive to irrigation. The relationship between GY under drought and irrigated treatments indicated that several RILs were superior to the best parent, SEA 5, and the four common bean check genotypes. Among the 118 genotypes tested, MR 8 was the most poorly adapted RIL under drought stress conditions.

A positive and significant correlation was observed between CB and GY under both irrigated and drought conditions with values of 0.52*** and 0.60***, respectively (Table 1). The genotype Cowpea cv. Mouride showed the highest value of CB under stress conditions (Figure 3). But this genotype showed a lower value of HI. Five lines (MR 112, MR 25, MR 93, MR 12, and MR 52) combined higher CB-values with higher GY-values under drought stress conditions (Figure 3). The line MR81 was outstanding in its grain yield under drought conditions, but its CB-value was not high under stress conditions. The susceptible check DOR 390 and five RILs (MR 8, EAP9510-77, MR 3, MR 42, and MR 116) showed poor adaptation to drought conditions with lower values of CB and GY under drought conditions (Figure 3). The drought adapted parent SEA 5 showed higher values of CB and GY under drought stress conditions than the susceptible parent MD 23-24 (Figure 3).

Poor correlation was observed between GY and PPI under drought conditions (Table 1). However, four RILs (MR 12, MR 77, MR 27, and MR 2) combined higher value of PPI and GY under drought stress conditions. Three RILs (MR 34, MR 17, and MR 119) were outstanding in mobilizing photosynthates to pod formation, but the CB-values of these lines were lower under drought stress, which resulted in lower values of GY. The PHI reflects the ability to mobilize photosynthates from pod wall to seed. A positive and highly significant correlation between PHI and GY under both irrigated and drought conditions was observed (Table 1). Nine genotypes (Cowpea, MR 81, MR 95, MR 120, MR 110, SEA 15, MR 52, MR 93, and MR 25) were superior in their ability to mobilize photosynthates from pod wall to grain, resulting in a higher grain yield under drought conditions (Figure 4). Two RILs (MR 112 and MR 12) showed poor performance in photosynthate mobilization to grain formation. Five genotypes (EAP 9510-77, MR 8, MR 116, MR 109, and DOR 390) combined low values of PHI with low values of GY under drought stress (Figure 4). The drought adapted parent SEA 5, with better value of GY than the susceptible parent MD 23-24, presented lower values of PHI than MD 23-24 under drought stress conditions, and slightly below the average of the RILs evaluated. Results on the relationship between the values of GY and HI under drought stress indicated that MR 25 and MR 81 were superior in mobilizing photosynthates to seeds. The HI-values of EAP 9510-77, MR 109 and MR 40 were markedly lower than that of other bean genotypes.

A negative and significant correlation (-0.31^{***}) was observed between DPM and GY under drought conditions (Table 1). Under irrigated conditions the DPM of 118 genotypes ranged from 62 to 70 days with a mean of 66 days; under drought stress the DPM ranged from 61 to 69 with a mean of 66 days (Data not shown). A total of 22 lines showed shorter and similar DPM under both irrigated and drought conditions (Data not shown). Another group of 22 lines showed shorter DPM with superior values of GY than the other genotypes under drought stress conditions (Data not shown). A negative and significant correlation ($r = -0.19^*$) was observed between DPM and canopy biomass under drought stress conditions. A positive and significant correlation was observed between DPM and SNA under irrigated and drought conditions ($r = 0.50^{***}$ and $r = 0.31^{***}$), respectively. The SNA showed a positive and highly significant correlation with grain yield under both irrigated and drought treatments (Table 1). Ten genotypes (Cowpea cv. Mouride, MR 102, MR 81, MR 114, MR 93, MR 77, MR 35, MR 25, MR 117, and MR 65) showed higher values of SNA than the other genotypes under drought stress conditions (Figure 5). Six genotypes (MR 66, MR 92, MR 26, EAP 9510-77, MR 8, and MR 40) were characterized by low values of SNA under drought stress. The parent MD 23-24 presented higher SNA under drought conditions than the parent SEA 5 (Figure 5), but SEA 5 was superior in its 100 seed weight (SW).

A positive and significant correlation ($r = 0.52^{***}$) was observed between SNA and PHI under drought conditions. Five lines (MR 81, MR 110, MR 95, MR 93, MR 120) combined higher

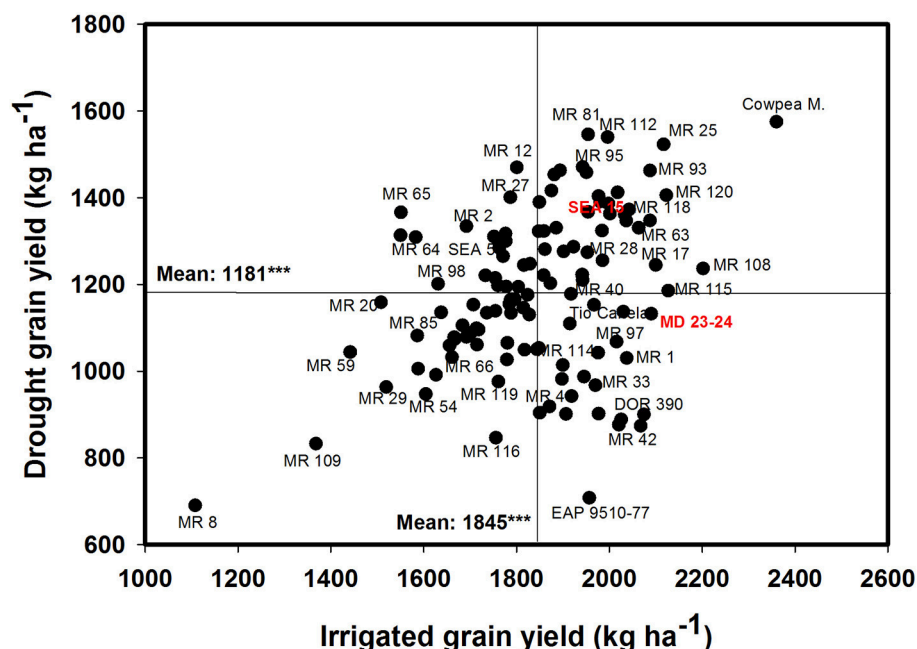


FIGURE 2 | Identification of genotypes that are adapted to drought stress conditions and are responsive to irrigation in a Mollisol at Palmira. Genotypes that yielded superior with drought and were also responsive to irrigation were identified in the upper, right hand quadrant.

SNA with higher PHI values under drought stress conditions (Figure 6). These lines also presented higher grain yield under drought stress (Figure 2). The susceptible check DOR 390, the parent MD 23–24 and five RILs (MR 114, MR 12, MR 11, MR 35) showed higher SNA combined with lower PHI and GY under drought conditions (Figures 2, 6). The lines EAP 9510-77, MR 116, MR 88, MR 49, and MR 8 showed poor adaptation to drought conditions with lower values of SNA, PHI, and GY under drought conditions (Figures 2, 6).

Phenotypic Evaluation of Root Traits in Soil Cylinder System

Drought stress reduced root growth in terms of total root length (TRL) compared to irrigated conditions, from an average of 45.9 m plant⁻¹ under irrigated conditions to 39.4 m plant⁻¹ under drought stress. However, it is noteworthy that several RILs (MR22, MR67, MR87, MR76, MR36, MR104, MR42, MR86, MR107, MR45, MR83, MR85, MR1, MR7, MR80, MR113, MR40, MR66, MR17, MR23, MR33, and MR58) and parents (SEA 5 and MD23–24) were characterized by increased root production under drought stress compared with irrigated conditions (Data not shown). Also an increase in production of deep roots (root length at soil depth 60–75 cm, m plant⁻¹) under drought stress was observed for the above mentioned lines (Table 2). Relationship between deep rooting evaluated under greenhouse conditions and GY in field conditions was tested and it didn't show a significant correlation under both irrigated and drought conditions (Table 1).

A significant negative correlation ($r = -0.22^*$) was observed between roots produced at soil depth 0–5 cm with GY

under drought stress conditions (Table 1). Also a positive and significant correlation ($r = 0.21^*$) was observed between roots length production at soil depth 20–40 cm with GY under drought stress (Table 1). A wide range of diversity in TRL was observed under drought conditions (Figure 7). Some genotypes such as one of the parents (SEA 5) and six RILs (MR 81, MR 12, MR 93, MR 25, MR 52, and MR 67) combined vigorous root system with superior values of GY under drought stress, while the three drought sensitive checks (Tio Canela, DOR 390, EAP 9510-77) and five RILs (MR 116, MR 54, MR 78, MR 109, and MR 69) showed poor root growth with lower values of GY under drought stress (Figure 7). Two genotypes (Cowpea, SEA 15) and two RILs (MR 112 and MR 120) were outstanding in their GY under drought stress but had poorer root growth compared with the other lines tested. Contrary to this observation, one parent (MD 23–24) and four RILs (MR 8, MR 3, MR 49, and MR 29) had vigorous root growth but lower GY under drought stress (Figure 7). A strong and positive relationship ($r = 0.68^{***}$) between vigorous root growth in term of greater value of TRL, and deep rooting ability in terms of root production at soil depth of 60–75 cm was observed under drought stress. The drought resistant parent SEA 5 was outstanding in its deep rooting ability under drought stress (Figure 8). Several RILs combined deep rooting ability with higher GY under drought stress such as MR 25, MR 93, MR 67, MR 81, MR 95, MR 12, and MR 32 (Figure 8). Three RILs (MR 112, MR 24 and MR 120) showed greater GY and shallow root development under drought stress. Five lines (MR 13, MR 31, MR 49, MR 3, and MR 22) were identified as outstanding in their deep rooting ability under drought stress, but not in producing GY (Figure 8).

TABLE 1 | Correlation coefficients (r) between final grain yield (GY) and other plant attributes and canopy biomass (CB) and other plant attributes of RILs of common bean grown under irrigated and drought stress conditions in a Mollisol in Palmira, Colombia.

Plant traits	Canopy biomass (kg ha ⁻¹)		Grain yield (kg ha ⁻¹)	
	Irrigated	Drought	Irrigated	Drought
SPAD chlorophyll meter readings (SCMR)	0.40***	0.03	0.35***	0.06
Leaf area index (LAI; m ² m ⁻²)	0.68***	0.56***	0.28**	0.30**
Canopy biomass (CB; kg ha ⁻¹)	1.00	1.00	0.52***	0.60***
Pod harvest index (PHI; %)	0.06	0.20*	0.44***	0.49***
Pod partitioning index (PPI; %)	-0.11	0.18*	0.15	0.09
Harvest index (HI; %)	-0.38***	-0.34***	0.05	0.13
100 seed weight (SW; g)	-0.13	0.14	-0.04	0.22*
Grain yield (GY; kg ha ⁻¹)	0.52***	0.60***	1.00	1.00
Shoot TNC (mg g ⁻¹)	0.07	0.22*	0.32***	0.20*
Seed TNC (mg g ⁻¹)	-0.09	0.19*	0.08	0.20*
Pod number per area (PNA)	0.45***	0.25**	0.29**	0.03
Seed number per area (SNA)	0.64***	0.38***	0.40***	0.34***
Days to maturity (DM)	0.22*	-0.20*	0.09	-0.31***
Visual rooting depth (VRD) at flowering (cm plant ⁻¹)	0.02	0.18	-0.02	0.18
Root length at soil depth 0–5 cm (RL0–5; m plant ⁻¹)	-0.07	-0.34***	-0.11	-0.22*
Root length at soil depth 5–10 cm (RL5–10; m plant ⁻¹)	-0.05	-0.03	-0.22*	0.05
Root length at soil depth 10–20 cm (RL10–20; m plant ⁻¹)	0.03	0.11	-0.08	0.17
Root length at soil depth 20–40 cm (RL20–40; m plant ⁻¹)	0.09	0.11	-0.01	0.21*
Root length at soil depth 40–60 cm (RL40–60; m plant ⁻¹)	0.11	0.04	-0.02	0.12
Root length at soil depth 60–75 cm (RL60–75; m plant ⁻¹)	0.16	0.15	0.01	0.11
Total root length (TRL; m plant ⁻¹)	0.09	0.00	-0.07	0.11
Fine root proportion at soil depth 0–5 cm (FRP0–5; %)	-0.18*	-0.47***	-0.18	-0.27**
Fine root proportion at soil depth 5–10 cm (FRP5–10; %)	-0.21*	-0.35***	-0.14	-0.11
Fine root proportion at soil depth 10–20 cm (FRP10–20; %)	-0.18	-0.27**	-0.06	-0.08
Fine root proportion at soil depth 20–40 cm (FRP20–40; %)	-0.14	-0.22*	-0.04	-0.07
Fine root proportion at soil depth 40–60 cm (FRP40–60; %)	0.03	-0.06	0.22*	0.08
Fine root proportion at soil depth 60–75 cm (FRP60–75; %)	0.05	0.13	0.03	0.11
Total fine root proportion (TFRP; %)	0.00	-0.04	0.08	0.07

*, **, *** Significant at the 0.05, 0.01, and 0.001 probability levels, respectively.

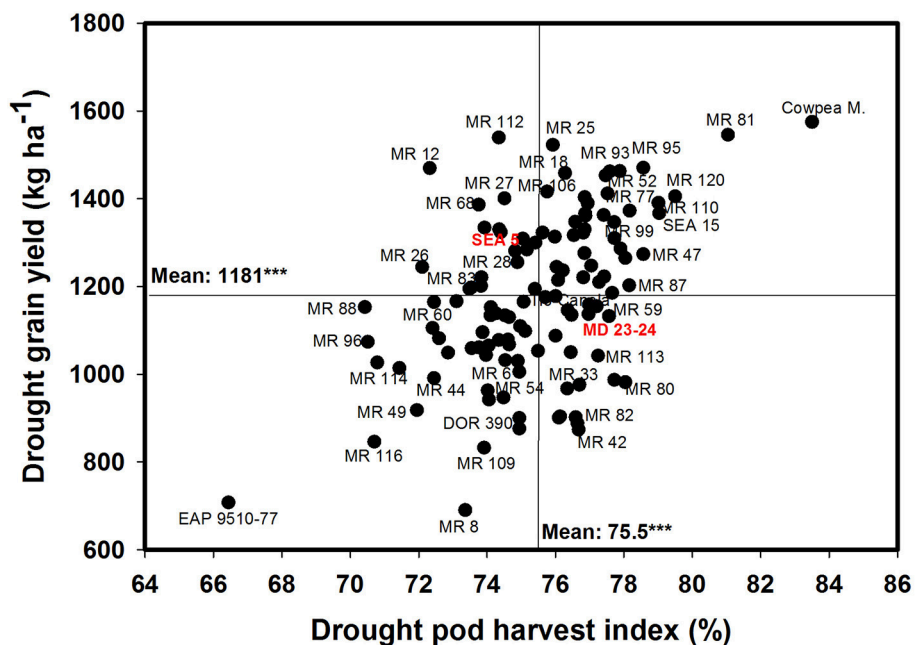
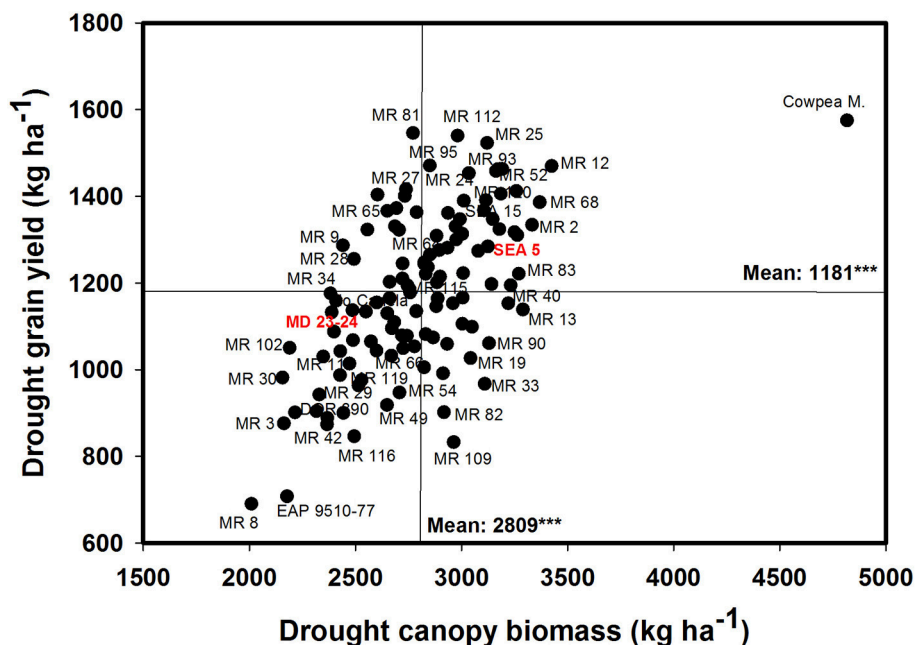
A slight increase in the production of fine roots was observed under drought stress compared to irrigation, from an average in fine root proportion of 81% under irrigation to 83% under drought (Table 2). Several RILs were superior to the parents (SEA 5 and MD 23–24) in the increase of fine root production under drought stress. Both parents showed almost similar fine root proportion under both irrigated and drought conditions and were superior to the average value of the RIL population under both conditions (Table 2). No clear relationship between fine root development and grain yield was observed under both irrigated and drought conditions; but under drought stress a significant negative correlation was observed between fine root proportion at different soil layers (0–5, 5–10, 10–20, and 20–40) with CB.

Nineteen RILs (MR11, MR13, MR15, MR19, MR25, MR51, MR6, MR67, MR68, MR79, MR9, MR93, MR98, MR90, MR106, MR3, MR97, MR40, and MR52), and the two parents (SEA 5 and MD 23–24) showed rapid root growth under drought stress, as reflected by the maximum visual rooting depth values under drought conditions (Data not shown). The lower yielding genotypes under drought stress such as DOR 390 and Tio

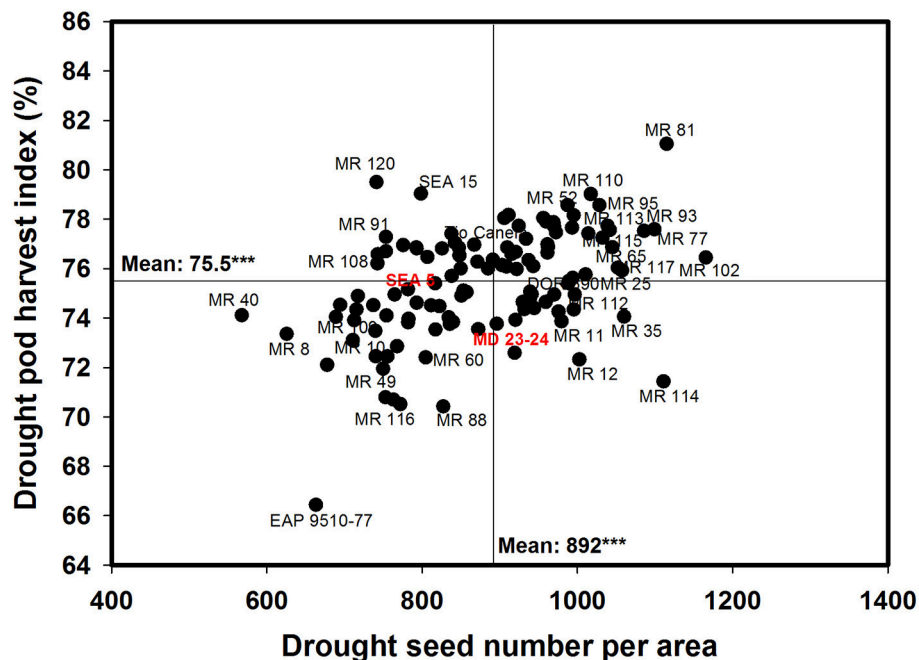
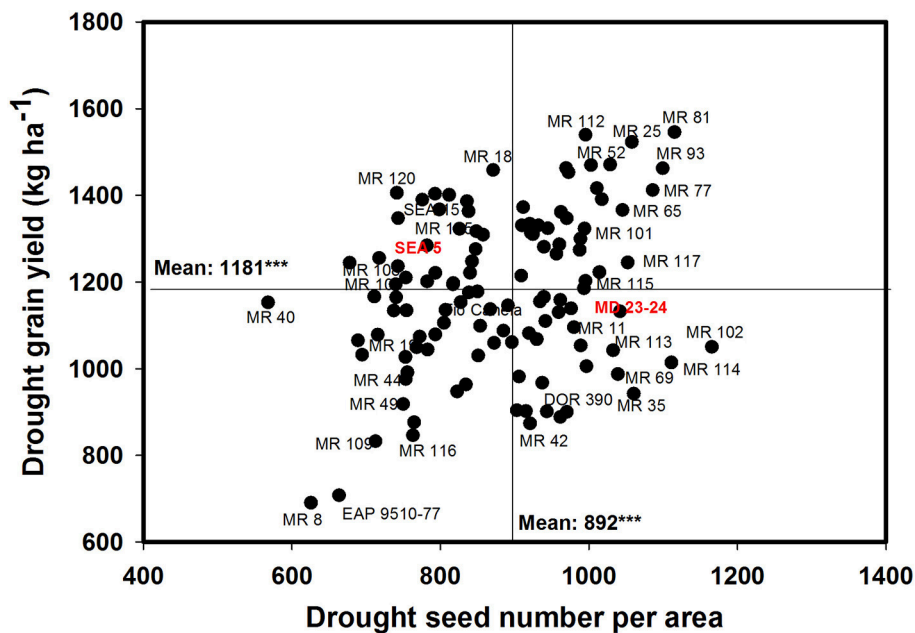
Canela 75 (drought sensitive and commercial checks) were characterized by slow root growth under drought conditions based on the lower values of visual rooting depth. Also these two checks presented a poor root development under drought and well-watered conditions in terms of TRL based on lower values of deep root production (root length at soil depth 60–75 cm, m plant⁻¹) and thicker roots under drought stress (Table 2; Figures 7, 8).

Principal Component Analysis

Under irrigated and drought stress conditions, eight principal components with cumulative variance of 75% was extracted which gives the clear idea of structure underlying the variables analyzed. Under irrigated conditions for Component 1 which has the contribution of visual rooting depth at flowering, root length at soil depth of 40–60 and 60–75 cm, TRL and fine root proportion for 21% of the total variability. For component 2, SCMR, LAI, CB, SW, GY, PNA, SNA, root length at soil depth of 0–5 cm, root length at soil depth of 10–20 cm, and TRL has contributed to 16% of total variability (Table 3). The



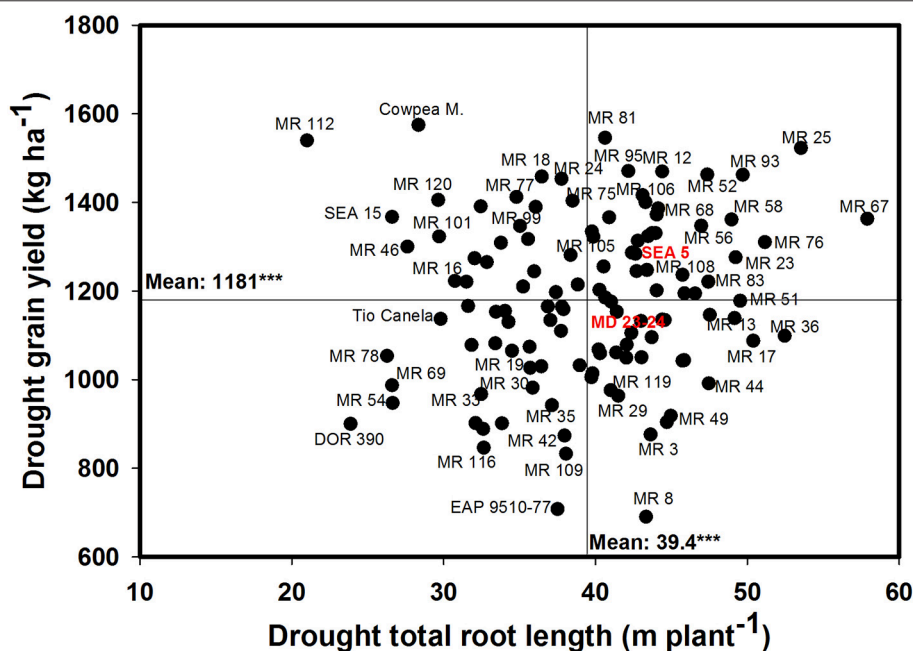
Under drought stress conditions for Component 1 which has the contribution of about 22% of the total variability from root traits such as, visual rooting depth at flowering, root length at soil depth of 10–20, 20–40, 40–60, and 60–75 cm, TRL, less fine roots at shallow soil layers and higher fine root proportion at deeper



showed that while deep rooting and earliness has contributed to superior performance under drought conditions, the formation of pods and seeds were not the factor limiting the grain yield. It

TABLE 2 | Phenotypic variation in root traits of parents and recombinant inbred lines (RILs) of MD23–24 × SEA 5 grown in soil cylinders under irrigated (well-watered) and drought stress conditions in the greenhouse at Palmira, Colombia.

Trait	Irrigated					Drought				
	Parents		RILs			Parents		RILs		
	MD 23–24	SEA 5	Max	Min	Mean	MD 23–24	SEA 5	Max	Min	Mean
Visual rooting depth at flowering (cm plant ⁻¹)	75	75	75	47	65	73	73	75	44	65
Root length at soil depth 60–75 cm (m plant ⁻¹)	2.1	2.6	9.0	0.0	3.1	2.7	5.2	6.0	0.0	2.2
Total root length (m plant ⁻¹)	34	37	78	23	46	43	43	58	21	40
Total fine root proportion (%)	87	84	89	68	81	88	86	90	68	83

**FIGURE 7 | Relationship between drought grain yield (GY) under field conditions and total root length (TRL) production under drought stress in greenhouse conditions.** Genotypes with higher values of GY and TRL under drought stress were identified in the upper, right hand quadrant.

was rather the ability to fill seeds as reflected by the significant positive associations between GY and PHI.

DISCUSSION

This study permitted evaluating shoot and root traits related with drought resistance in a set of 111 RILs of common bean developed for improving drought resistance. Since the study was conducted over three seasons with intermittent drought stress (occurring on and off but especially around the vegetative to early reproductive period of plant development) it facilitated identification of few RILs with superior shoot traits that contributed to improved drought resistance. We complemented the field studies with a greenhouse study on root traits so that we can evaluate the role of root traits in combination with shoot traits for improving drought resistance. Previous research showed that bean genotypes derived from Durango race such as SEA 5 and SEA 15, have mechanisms that can maintain

a competitive level of water balance, allowing these genotypes to promote grain formation and filling during drought stress (Rosales et al., 2012; Beebe et al., 2013). By using a set of RILs we could dissect the physiological basis of the superior performance of lines improved for drought resistance. We found significant transgressive segregation for GY and several morpho-physiological shoot and root traits under both irrigated and drought stress conditions. The population distributions were continuous indicating quantitative inheritance for the traits measured.

Grain Yield and Canopy Biomass

Several shoot traits evaluated in this study showed transgressive segregation in both directions under drought stress, such as GY, CB, SNA, and PHI. Production of CB can be an indicator of the success of the plant in its net fixation of CO₂, assimilation of nutrients and effective use of water under both optimal and stress conditions, where a higher accumulation of assimilates is

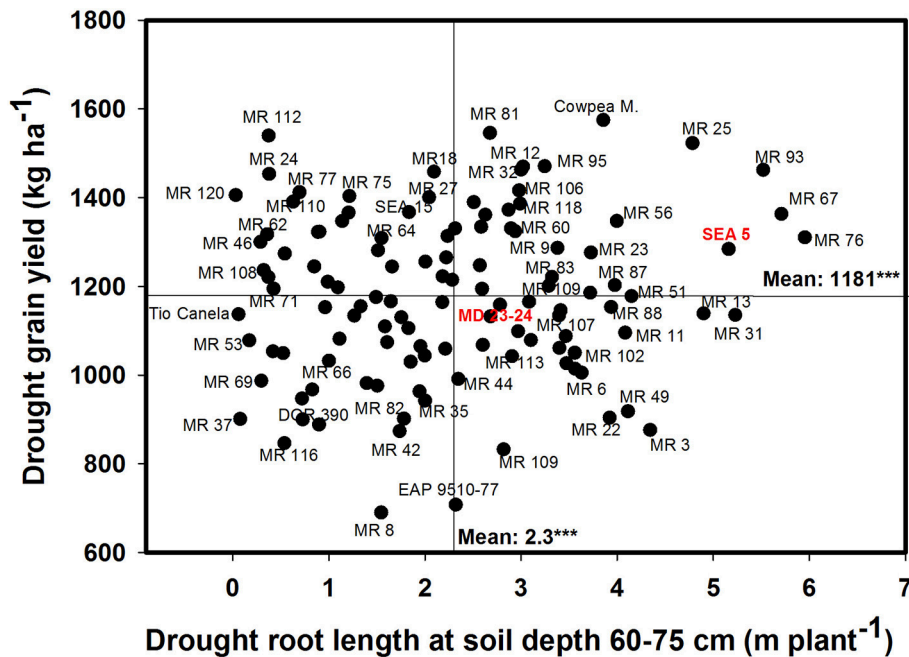


TABLE 3 | Eigen values and percent of total variation and component matrix for the principal component axes—Irrigated conditions.

Principal components	1	2	3	4	5	6	7	8
Eigen values	5.87	4.49	2.70	2.28	1.85	1.46	1.24	1.15
Proportion of variance	0.21	0.16	0.10	0.08	0.07	0.05	0.04	0.04
Cumulative	0.21	0.37	0.47	0.55	0.61	0.67	0.71	0.75
COMPONENT MATRIX								
SCMR	0.094	-0.239	0.096	0.156	-0.041	-0.288	0.020	-0.037
LAI	0.177	-0.247	0.243	-0.062	0.116	-0.163	-0.087	0.247
CB	0.135	-0.265	0.167	0.083	0.388	-0.140	0.086	0.107
PHI	-0.021	-0.118	-0.096	0.373	-0.015	0.316	0.153	-0.053
PPI	-0.078	0.048	-0.056	0.126	0.065	0.289	-0.467	0.456
HI	-0.014	0.000	-0.108	0.192	-0.486	0.209	-0.066	0.146
SW	-0.032	0.288	-0.182	-0.002	0.309	-0.155	0.014	0.263
GY	0.051	-0.214	-0.023	0.385	0.281	0.071	-0.039	0.210
Shoot TNC	-0.107	0.001	-0.151	0.282	0.405	0.051	0.114	-0.267
Seed TNC	-0.061	-0.023	-0.141	0.193	-0.136	0.363	0.394	0.118
PNA	0.090	-0.339	0.192	0.065	-0.139	0.100	0.101	-0.053
SNA	0.141	-0.371	0.159	0.107	0.000	0.082	0.146	-0.104
DM	0.159	-0.210	0.254	-0.081	-0.268	0.001	-0.217	0.136
VRD flowering	0.319	0.097	-0.205	0.065	-0.128	-0.157	0.047	-0.068
RL0-5	0.029	0.223	0.311	0.039	0.047	-0.049	0.193	0.334
RL5-10	0.093	0.162	0.082	-0.114	-0.039	-0.093	0.551	0.113
RL10-20	0.204	0.266	0.260	0.079	0.104	0.219	0.001	0.070
RL20-40	0.294	0.178	0.148	0.063	0.097	0.253	-0.111	-0.162
RL40-60	0.333	0.103	0.022	-0.051	0.066	0.201	-0.071	-0.242
RL60-75	0.324	0.032	-0.035	-0.155	0.068	0.162	-0.066	-0.189
TRL	0.310	0.204	0.184	-0.004	0.077	0.168	-0.029	-0.055
FRP0-5	-0.128	0.197	0.385	0.091	-0.095	0.006	0.164	0.123
FRP5-10	-0.200	0.134	0.284	0.192	-0.130	-0.059	0.087	-0.173
FRP10-20	-0.213	0.159	0.316	0.199	-0.006	-0.063	-0.190	-0.140
FRP20-40	-0.174	0.100	0.163	0.287	-0.021	-0.173	-0.228	-0.351
FRP40-60	0.152	0.089	-0.140	0.361	-0.020	-0.164	-0.004	0.122
FRP60-75	0.304	0.077	-0.153	0.148	-0.179	-0.254	-0.021	0.041
TFRP	0.255	0.147	-0.089	0.332	-0.178	-0.304	-0.037	0.033

SCMR, SPAD chlorophyll meter readings; LAI, Leaf area index ($m^2 m^{-2}$); CB, Canopy biomass ($kg ha^{-1}$); PHI, Pod harvest index (%); PPI, Pod partitioning index (%); HI, Harvest index (%); SW, 100 seed weight (g); GY, Grain yield ($kg ha^{-1}$); Shoot TNC, Shoot TNC ($mg g^{-1}$); Seed TNC, Seed TNC ($mg g^{-1}$); PNA, Pod number per m^2 ; SNA, Seed number per m^2 ; DM, Days to maturity; VRD Flowering, Visual rooting depth at flowering ($cm plant^{-1}$); RL0-5, Root length at soil depth 0-5 cm ($m plant^{-1}$); RL5-10, Root length at soil depth 5-10 cm ($m plant^{-1}$); RL10-20, Root length at soil depth 10-20 cm ($m plant^{-1}$); RL20-40, Root length at soil depth 20-40 cm ($m plant^{-1}$); RL40-60, Root length at soil depth 40-60 cm ($m plant^{-1}$); RL60-75, Root length at soil depth 60-75 cm ($m plant^{-1}$); TRL, Total root length ($m plant^{-1}$); FRP0-5, Fine root proportion at soil depth 0-5 cm (%); FRP5-10, Fine root proportion at soil depth 5-10 cm (%); FRP10-20, Fine root proportion at soil depth 10-20 cm (%); FRP20-40, Fine root proportion at soil depth 20-40 cm (%); FRP40-60, Fine root proportion at soil depth 40-60 cm (%); FRP60-75, Fine root proportion at soil depth 60-75 cm (%); TFRP, Total fine root proportion (%).

under irrigated and intermittent drought conditions, because of possible additional vegetative growth occurring after mid-pod filling to physiological maturity due to irrigation or rainfall. The distribution of rainfall in the 3 years of evaluation indicate that the crop suffered intermittent drought stress, some years the rainfall during grain filling stage was higher than the other years (Figure 1). This additional water availability can cause additional vegetative growth that is difficult to estimate (due to leaf fall during this period). We assume that the plant can take advantage of this additional water to improve grain formation and filling which could result in better grain filling under intermittent than terminal drought stress. The correlation coefficients in this study between GY and PPI were not positive

and significant, possibly due to the effect of additional rainfall, especially 2007 compared to the other 2 years. These conditions may make some lines to revert back to the behavior of wild bean (Beebe et al., 2008) exhibiting different patterns of growth and remobilization, making it unclear the major contribution of this trait. However, several RILs combined superior GY with PPI under intermittent drought stress indicating their superior ability to remobilize photosynthates from vegetative plant structures to pod production.

The contribution of remobilization of photoassimilates from vegetative structures to the pod and grain production for improving drought resistance has been reported either by estimating dry matter partitioning indices such as HI, PPI, and

TABLE 4 | Eigen values and percent of total variation and component matrix for the principal component axes—Drought stress.

Principal components	1	2	3	4	5	6	7	8
Eigen values	6.22	3.83	2.74	2.49	1.99	1.28	1.15	1.02
Proportion of variance	0.22	0.14	0.10	0.09	0.07	0.05	0.04	0.04
Cumulative	0.22	0.36	0.46	0.55	0.62	0.66	0.70	0.74
COMPONENT MATRIX								
SCMR	0.012	−0.061	0.201	0.052	−0.058	−0.461	−0.361	0.453
LAI	0.157	−0.161	−0.145	−0.024	0.252	0.354	−0.131	0.012
CB	0.131	−0.274	−0.338	0.107	0.021	0.076	−0.240	0.254
PHI	−0.015	−0.157	0.001	0.482	0.094	−0.083	0.020	−0.100
PPI	0.044	0.065	−0.244	0.040	−0.070	0.059	0.362	0.444
HI	−0.059	0.014	0.229	0.295	0.094	−0.313	0.238	−0.413
SW	0.029	0.159	−0.408	−0.102	−0.109	−0.193	−0.063	−0.301
GY	0.114	−0.147	−0.271	0.345	0.016	−0.197	−0.265	−0.127
Shoot TNC	−0.068	−0.038	−0.267	0.180	−0.075	−0.089	−0.005	0.123
Seed TNC	0.016	−0.098	−0.150	0.270	−0.215	0.015	0.474	−0.038
PNA	0.045	−0.331	0.179	0.107	0.071	0.059	0.298	0.192
SNA	0.055	−0.379	0.189	0.288	0.057	−0.004	−0.005	0.117
DM	0.027	−0.167	0.365	−0.142	0.184	0.238	0.075	0.022
VRD Flowering	0.334	0.026	0.083	0.006	−0.251	−0.094	−0.027	−0.016
RL0–5	−0.107	0.264	0.069	−0.021	0.062	−0.253	0.170	0.268
RL5–10	0.169	0.272	−0.103	0.091	0.221	−0.058	0.152	0.218
RL10–20	0.251	0.196	−0.132	0.067	0.302	0.038	0.158	0.024
RL20–40	0.298	0.149	−0.035	0.072	0.236	0.035	−0.018	−0.013
RL40–60	0.330	0.042	0.168	0.028	0.130	−0.051	−0.034	−0.129
RL60–75	0.312	0.018	0.136	0.008	0.104	−0.053	−0.188	−0.074
TRL	0.310	0.228	0.034	0.050	0.264	−0.089	0.068	0.052
FRP0–5	−0.201	0.273	0.187	0.127	0.020	−0.084	−0.098	0.168
FRP5–10	−0.197	0.253	0.088	0.273	0.038	0.148	−0.210	0.050
FRP10–20	−0.195	0.217	0.053	0.311	0.084	0.285	−0.133	−0.009
FRP20–40	−0.170	0.193	−0.026	0.261	0.003	0.335	−0.081	0.003
FRP40–60	0.191	0.108	0.129	0.115	−0.415	0.004	0.080	−0.006
FRP60–75	0.290	0.059	0.077	0.025	−0.312	0.204	−0.073	0.032
TFRP	0.234	0.173	0.141	0.156	−0.406	0.215	−0.064	0.038

SCMR, SPAD chlorophyll meter readings; LAI, Leaf area index ($m^2 m^{-2}$); CB, Canopy biomass ($kg ha^{-1}$); PHI, Pod harvest index (%); PPI, Pod partitioning index (%); HI, Harvest index (%); S100W, 100 seed weight (g); GY, Grain yield ($kg ha^{-1}$); Shoot TNC, Shoot TNC ($mg g^{-1}$); Seed TNC, Seed TNC ($mg g^{-1}$); PNA, Pod number per m^2 ; SNA, Seed number per m^2 ; DM, Days to maturity; VRD Flowering, Visual rooting depth at flowering ($cm plant^{-1}$); RL0–5, Root length at soil depth 0–5 cm ($m plant^{-1}$); RL5–10, Root length at soil depth 5–10 cm ($m plant^{-1}$); RL10–20, Root length at soil depth 10–20 cm ($m plant^{-1}$); RL20–40, Root length at soil depth 20–40 cm ($m plant^{-1}$); RL40–60, Root length at soil depth 40–60 cm ($m plant^{-1}$); RL60–75, Root length at soil depth 60–75 cm ($m plant^{-1}$); TRL, Total root length ($m plant^{-1}$); FRP0–5, Fine root proportion at soil depth 0–5 cm (%); FRP5–10, Fine root proportion at soil depth 5–10 cm (%); FRP10–20, Fine root proportion at soil depth 10–20 cm (%); FRP20–40, Fine root proportion at soil depth 20–40 cm (%); FRP40–60, Fine root proportion at soil depth 40–60 cm (%); FRP60–75, Fine root proportion at soil depth 60–75 cm (%); TFRP, Total fine root proportion (%).

PHI (Hall, 2004; Rosales-Serna et al., 2004; Klaedtke et al., 2012; Rosales et al., 2012; Assefa et al., 2013; Rao et al., 2013, 2016a; Beebe et al., 2014; Rao, 2014; Polania et al., 2016a,c) or by quantifying starch and sugar accumulation and partitioning (Cuellar-Ortiz et al., 2008; Rosales et al., 2012; Andrade et al., 2016). Field evaluation in this study over three seasons under intermittent drought stress showed stronger correlation between PHI and GY confirming the contribution of mobilization of photosynthates from pod walls to grain (Table 1). It is important to point out that while several lines were superior in their PHI and GY, two RILs (MR 12 and MR 112) were high yielding under drought stress but presented thicker pod walls (relatively lower than average values of PHI; Figure 4). These two lines

had relatively higher values of CB and SNA that contributed to greater GY under drought stress. Improving the values of PPI and PHI in these two lines could improve further GY-values of these lines under drought stress. These two lines could further improve their GY-values under drought stress by enhancing their remobilization ability of photosynthates from pod wall to seed filling (i.e., increase in PHI). These results are consistent with previous reports which suggested that PHI could serve as a useful selection criteria for improving drought resistance in common bean because of its simplicity in measurement, significant correlation with GY under both irrigated and drought conditions and high heritability (Assefa et al., 2013; Rao et al., 2013; Beebe et al., 2014; Polania et al., 2016a).

Drought stress is known to reduce yield components such as PNA and SNA (Rao et al., 2013; Assefa et al., 2015). Seed number per pod has been identified as useful criteria for selection for improving drought resistance because of its higher heritability and contribution to genetic gain (Ramirez-Vallejo and Kelly, 1998). The decrease in the formation of pods and grains under drought stress is due to several factors, including pollen grain sterility that reduces pollen grain germination and pollen tube growth, and inadequate photosynthate supply that prevents embryo development (Farooq et al., 2016). Selection of genotypes that have greater sink strength reflected in greater values of SNA is required to increase GY under drought conditions. Our results demonstrate this relationship to be a positive and significant correlation between SNA and GY under drought stress. However, it is noteworthy that there are genotypes that have higher values of SNA but lower values of GY under drought stress (**Figure 5**), indicating that these genotypes are failing in their ability to optimize photosynthate mobilization to support grain filling process. This behavior can be evidenced in the relationship between SNA and PHI (**Figure 6**), in which genotypes such as MD 23–24, MR 114, MR 12, MR 11, and MR 35 showed higher SNA but lower PHI under drought conditions. Thus, these genotypes are capable of setting seeds but are poor in their ability to fill the seeds. An increase in photosynthate remobilization together with improved sink strength was observed in the superior genotypes that combined higher values of GY with SNA and PHI under drought stress conditions (**Figures 5, 6**).

Grain Yield and Physiological Efficiency through Early Maturity

The significant negative relationship between GY and DM under drought stress (**Table 1**) indicated that early maturing genotypes were more adapted to drought stress. The common bean farmers have multiple reasons for preferring short season varieties, an important one among them is to minimize exposure to drought (White and Singh, 1991; Beebe, 2012). Earliness is more useful where terminal drought predominates (Beebe et al., 2014) but a shorter growth cycle can reduce GY potential per day by an estimated value of 74 kg ha⁻¹ (White and Singh, 1991). This penalty in GY per day could be markedly reduced or even completely eliminated through improved physiological efficiency of the plant through genetically improving the capacity of the plant to produce more seeds and especially together with the ability to have better filling of these seed under drought stress (i.e., greater sink strength). In common bean different field studies showed that early maturing genotypes with superior photosynthate remobilization ability can yield better under both drought and irrigated conditions (Klaedtke et al., 2012; Rao et al., 2013, 2016a; Beebe et al., 2014; Polania et al., 2016a,c). This improved physiological efficiency under drought stress was shown to be independent of yield potential and phenological plasticity (Polania et al., 2016c). The phenotypic correlations between GY and CB, PHI, DPM, SNA, and SW suggest that in common bean higher values of CB combined with an efficient remobilization of photosynthates to the pod and grain formation could contribute to greater sink strength through higher values of both SNA and SW.

Grain Yield and Root Traits

Root growth and shoot growth have a complex relationship. In general, the shoot provides the root with carbon and certain hormones, and the root provides the shoot with water, nutrients, and also with hormones. To increase grain yield through a better plant growth under both optimal and drought stress conditions, the root system must be able to supply water and nutrients to the new plant growth without sequestering too much photoassimilates from the shoot (Bingham, 2001). Defining the morpho-physiological traits and mechanisms that are suited to different agroecological niches will play an important role in the development of new varieties adapted to different types of drought stress. Different studies on common bean and other crops contributed to define the root system characteristics for superior resistance to drought (Lynch, 2013). Among the root characteristics evaluated in this study, none stood out for its outstanding contribution or correlation with more grain production under drought stress and even under irrigated conditions (**Table 1**). This indicates the complexity of the relationship between root system development and drought resistance.

Results from this study demonstrated transgressive segregation in both directions under drought stress in several root traits evaluated, such as total root length production, deep root production, visual rooting depth, and fine root proportion. Several RILs exhibited markedly superior or lower expression of root traits than both parents (SEA 5 and MD 23–24) under drought as well as irrigated conditions. Transgressive segregation in root traits in common beans under irrigated and drought conditions have been observed also in RILs population of BAT 477 × DOR 364 (Asfaw and Blair, 2012). It was observed in several drought resistant RILs that root production is stimulated by drought stress compared with irrigated conditions and this could be an adaptive response of the plant to drought stress (Turner, 1979; Rao, 2014).

Results on phenotypic correlations and the multivariate analysis using the data on different shoot and root traits indicated that improved resistance to intermittent drought stress in common bean could result from different plant strategies. These strategies include different combinations of shoot and root traits that allows the plant to better adapt to drought stress. Based on the phenotypic differences in GY, CB, VRD and TRL, we classified the drought resistant lines into two groups, water savers and water spenders (**Table 5**). The water spender type genotypes combined higher values of GY with SNA under drought stress with rapid development of deep root system that allows the plant a faster access to available water in deep soil profiles (**Table 5**). This response to drought allows to continue the processes of gas exchange and carbon accumulation and facilitates improved remobilization of photosynthates resulting in an increased values of SNA and GY under drought stress. This overall response at whole plant level reflects an EUW rather than improved WUE resulting from partial closure of stomata (Polania et al., 2016a). Results from water spender type genotypes indicate that a deeper and vigorous root system helps to the plant to support a better sink strength that was reflected in higher values of SNA under drought conditions (**Table 5**). These water spender type genotypes would be better suited to

TABLE 5 | Combination of shoot and root traits of bean lines classified such as water saver type, water spender type and drought susceptible type that were grown under drought stress under greenhouse and field conditions at CIAT-Palmira, Colombia.

Genotype	Grain yield (kg ha ⁻¹)		Canopy biomass (kg ha ⁻¹)		Pod harvest index (%)		Pod partitioning index (%)		Seed number per area (m ²)		Visual rooting depth at flowering (cm plant ⁻¹)		Root length at soil depth 60–75 cm (m plant ⁻¹)		Total root length (m plant ⁻¹)	
	Irrigated	Drought	Irrigated	Drought	Irrigated	Drought	Irrigated	Drought	Irrigated	Drought	Irrigated	Drought	Irrigated	Drought	Irrigated	Drought
DROUGHT RESISTANT LINES WITH DEEP ROOTING ABILITY (WATER SPENDER TYPE)																
MR 81	1,954	1,546	4,433	2,771	83	81	54	47	1,277	1,115	69	63	2.3	2.7	52	41
MR 25	2,117	1,523	4,665	3,121	79	76	71	53	1,187	1,058	75	75	5.9	4.8	55	54
MR 95	1,943	1,471	4,865	2,850	79	79	45	53	1,384	1,029	69	70	4.1	3.2	50	42
MR 93	2,088	1,462	4,213	3,172	79	78	57	57	1,217	1,099	72	75	4.8	5.5	52	50
MR 67	2,001	1,363	4,464	2,788	79	77	59	69	1,286	838	68	75	6.7	5.7	39	58
SEA 5	1,763	1,284	4,787	3,124	77	75	66	54	1,042	782	75	73	2.6	5.2	37	43
DROUGHT RESISTANT LINES WITHOUT DEEP ROOTING ABILITY (WATER SAVER TYPE)																
MR 112	1,996	1,539	4,302	2,982	77	74	78	59	1,271	996	58	58	0.3	0.4	35	21
MR 24	1,881	1,453	4,864	3,034	80	77	67	55	1,124	973	56	55	2.1	0.4	47	38
MR 77	2,018	1,412	4,509	3,259	80	78	67	73	1,256	1,086	69	66	2.0	0.7	34	35
MR 120	2,123	1,405	5,086	3,186	82	80	61	57	1,006	741	60	57	0.4	0.0	37	30
MR 75	1,977	1,403	4,395	2,604	80	77	66	62	1,143	793	64	63	0.6	1.2	44	39
MR 110	1,980	1,391	4,653	3,115	81	79	57	73	1,123	1,017	71	62	2.8	0.6	54	32
DROUGHT SUSCEPTIBLE TYPE GENOTYPES																
MR 69	1,945	987	4,021	2,427	81	78	55	58	1,297	1,039	56	54	0.8	0.3	33	27
MR 54	1,604	947	3,710	2,707	79	74	70	71	1,133	822	55	55	1.7	0.7	34	27
DOR 390	2,075	900	4,593	2,443	81	75	70	62	1,432	970	68	54	2.6	0.7	33	24
MR 116	1,755	846	3,500	2,494	80	71	63	53	1,332	763	66	70	0.7	0.5	49	33
MR 109	1,368	833	3,478	2,963	79	74	46	58	1,017	713	74	60	2.9	2.8	44	38
EAP 9,510-77	1,957	708	4,558	2,177	80	66	51	51	1,259	663	48	65	0.1	2.3	27	38

intermittent drought conditions where water could be available at depth.

The strategy of water spender type genotypes under intermittent drought stress conditions may be that the deeper roots with better water extraction capacity can support the rate of photosynthesis and the accumulation of water soluble carbohydrates in the stem and this accumulated photosynthate could be remobilized to grain filling (Lopes and Reynolds, 2010). Several studies have demonstrated the contribution of deep rooting in increased water extraction from lower soil depth and its relationship with superior resistance to drought (Sponchiado et al., 1989; White and Castillo, 1992; Lynch and Ho, 2005; Ryser, 2006; Polania et al., 2009, 2012; Asfaw and Blair, 2012; Beebe et al., 2013; Rao, 2014; Rao et al., 2016b). Some efforts have been made in identifying genes and molecular markers that are associated with deep rooting ability (Asfaw and Blair, 2012) where a RILs population of DOR 364 × BAT 477 was used to identify quantitative trait loci (QTLs) that were associated with root traits under drought stress. Several QTLs were identified on linkage groups b01 or b11, which explained up to 41% of genetic variance.

The water saver type genotypes were superior in GY with moderate values of SNA under drought stress but their root system was slower in its development (Table 5). It is possible that these genotypes were better adapted to drought because of their ability to regulate stomatal opening for improved transpiration efficiency while maintaining their capacity to remobilize photosynthates toward pod and grain production (Polania et al., 2016a). These water saver type genotypes, would be suitable to semi-arid to arid environments where water availability is very limited with longer terminal drought stress conditions. Thus, further research on water use, photosynthesis and carbon mobilization is needed on the entire RIL population to classify the RILs as water savers or as water spenders and also to identify QTLs for different water use patterns based on both shoot and root traits.

Shallower rooting ability under drought stress can be complemented with traits related with conserving water at vegetative stage, such as lower leaf conductance, smaller leaf canopy, that would make more water available for reproductive growth and grain filling, resulting in better GY under terminal drought stress conditions (Zaman-Allah et al., 2011; Araújo et al., 2015). A poor root system can be limiting for an optimal plant development and grain production under drought stress, even for a good response in optimal conditions; as can be evidenced in the drought sensitive genotypes (Table 5). These genotypes showed less resistance to drought (lower values of GY) and were characterized by lower values of total root production as well as less proportion of roots at deeper soil layers under both irrigated and drought conditions. However, some RILs had deep root system but not higher values of GY thus indicating that deep rooting alone with lower sink strength will not result in higher values of GY under drought.

The results from both the field and greenhouse studies indicated the need to determine what size and what kind of distribution of root system is required in a certain field site to avoid a trade-off or any restriction in the plant growth and yield (Bingham, 2001). A very vigorous root system, in

an inefficient plant to assimilate CO₂, becomes plant's another sink, competing for photoassimilates with the economic organ of interest of the plant, and increase the sensitivity to drought stress. Thus, a vigorous and deeper root system, with rapid growth is useful but not enough to have greater resistance to drought in common bean. It is the strategic combination of traits that improves physiological efficiency such as a better developed root system helping the plant to access water to maintain transpiration rates and vegetative growth combined with the ability to remobilize photosynthates from vegetative structures to the pods and subsequently to grain production is what is needed for improved drought resistance (Beebe et al., 2014; Rao, 2014; Araújo et al., 2015; Polania et al., 2016a; Rao et al., 2016a).

CONCLUSIONS

We evaluated different root and shoot traits in a large RILs population and identified a few relevant traits for improved resistance to intermittent drought in common bean. The phenotypic data generated from this work will be useful to identify shoot and root QTLs associated with improved resistance to intermittent drought. Previous studies have reported the contribution of individual traits such as deep rooting, CB, water use, HI and PHI to the adaptation to drought stress in common bean, but not the combination of these traits and how the combination particularly with a focus on sink strength contributes to improved adaptation to intermittent drought stress. Our results indicate that common bean genotypes respond to drought stress as water spending types or water saver types. The water saver type of genotypes respond to drought with intermediate to shallow rooting system, high water use efficiency, reduced sink strength, and superior photosynthate remobilization to pod and grain formation. The water spender type of genotypes respond to drought with a better developed root system helping the plant to access water to maintain transpiration rates and vegetative growth, combined with the ability to remobilize photosynthates from vegetative structures to the pods and subsequently to seed production resulting in a superior number of pods and seeds per area. We observed transgressive segregation in root traits such as total root length and deep rooting ability under irrigated and drought conditions. We identified five RILs (MR 25, MR 93, MR 67, MR 81, MR 95) as drought resistant-water spender types and five RILs (MR 112, MR 24, MR 77, MR 120, MR 75) as drought resistant-water saver types. We identified rooting depth, CB, PPI, PHI, PNA, and SNA as useful traits for improving resistance to intermittent drought. Some of these traits are easier to implement in a breeding program due to their simplicity and relatively low analytical cost such as PHI.

AUTHOR CONTRIBUTIONS

JP, CC, SB, and IR designed the experiments and contributed to data interpretation. JP, CC, MG, MR, and FV collected and analyzed the data. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors acknowledge the support of the German Government through a restricted core grant from BMZ-GTZ Project No. 2002.7860.6-001.00 with Contract No. 81084613, and the CGIAR research program on grain

legumes for financial support of research on improving drought resistance in common bean. We would also like to thank all donors who supported this work through their contributions to the CGIAR Fund. We also thank bean breeding and physiology teams at CIAT, Colombia for their contribution.

REFERENCES

- Andrade, E. R., Ribeiro, V. N., Azevedo, C. V., Chiorato, A. F., Williams, T. C., and Carbonell, S. A. (2016). Biochemical indicators of drought tolerance in the common bean (*Phaseolus vulgaris* L.). *Euphytica* 210, 277–289. doi: 10.1007/s10681-016-1720-4
- Araújo, S. S., Beebe, S., Crespi, M., Delbreil, B., Gonzalez, E. M., Gruber, V., et al. (2015). Abiotic stress responses in legumes: strategies used to cope with environmental challenges. *CRC Crit. Rev. Plant Sci.* 34, 237–280. doi: 10.1080/07352689.2014.898450
- Araus, J. L., Slafer, G. A., Reynolds, M. P., and Royo, C. (2002). Plant breeding and drought in C3 cereals: what should we breed for? *Ann. Bot.* 89, 925–940. doi: 10.1093/aob/mcf049
- Asfaw, A., and Blair, M. W. (2012). Quantitative trait loci for rooting pattern traits of common beans grown under drought stress versus non-stress conditions. *Mol. Breed.* 30, 681–695. doi: 10.1007/s11032-011-9654-y
- Assefa, T., Beebe, S., Rao, I. M., Cuasquer, J., Duque, M. C., Rivera, M., et al. (2013). Pod harvest index as a selection criterion to improve drought resistance in white pea bean. *Field Crop. Res.* 148, 24–33. doi: 10.1016/j.fcr.2013.04.008
- Assefa, T., Wu, J., Beebe, S., Rao, I. M., Marcomin, D., and Claude, R. J. (2015). Improving adaptation to drought stress in small red common bean: phenotypic differences and predicted genotypic effects on grain yield, yield components and harvest index. *Euphytica* 303, 477–489. doi: 10.1007/s10681-014-1242-x
- Beebe, S. (2012). Common bean breeding in the tropics. *Plant Breed. Rev.* 36, 357–426. doi: 10.1002/9781118358566
- Beebe, S. E., Rao, I. M., Blair, M. W., and Acosta-Gallegos, J. A. (2013). Phenotyping common beans for adaptation to drought. *Front. Physiol.* 4:35. doi: 10.3389/fphys.2013.00035
- Beebe, S., Rao, I. M., Cajiao, C., and Grajales, M. (2008). Selection for drought resistance in common bean also improves yield in phosphorus limited and favorable environments. *Crop Sci.* 48, 582–592. doi: 10.2135/cropsci2007.07.0404
- Beebe, S., Rao, I. M., Devi, M., and Polania, J. (2014). Common beans, biodiversity, and multiple stress: challenges of drought resistance in tropical soils. *Crop Pasture Sci.* 65, 667–675. doi: 10.1071/CP13303
- Bingham, I. J. (2001). Soil root canopy interactions. *Ann. Appl. Biol.* 138, 243–251. doi: 10.1111/j.1744-7348.2001.tb00108.x
- Blum, A. (2009). Effective use of water (EUW) and not water-use efficiency (WUE) is the target of crop yield improvement under drought stress. *Field Crop. Res.* 112, 119–123. doi: 10.1016/j.fcr.2009.03.009
- Blum, A. (2015). Towards a conceptual ABA ideotype in plant breeding for water limited environments. *Funct. Plant Biol.* 42, 502–513. doi: 10.1071/FP14334
- Butare, L., Rao, I. M., Lepoivre, P., Polania, J., Cajiao, C., Cuasquer, J., et al. (2011). New genetic sources of resistance in the genus *Phaseolus* to individual and combined aluminium toxicity and progressive soil drying stresses. *Euphytica* 181, 385–404. doi: 10.1007/s10681-011-0468-0
- Condon, A. G., Richards, R. A., Rebetzke, G. J., and Farquhar, G. D. (2004). Breeding for high water-use efficiency. *J. Exp. Bot.* 55, 2447–2460. doi: 10.1093/jxb/erh277
- Cuellar-Ortiz, S. M., De La Paz Arrieta-Montiel, M., Acosta-Gallegos, J. A., and Covarrubias, A. A. (2008). Relationship between carbohydrate partitioning and drought resistance in common bean. *Plant Cell Environ.* 31, 1399–1409. doi: 10.1111/j.1365-3040.2008.01853.x
- Daryanto, S., Wang, L., and Jancinthe, P.-A. (2015). Global synthesis of drought effects on food legume production. *PLoS ONE* 10:e0127401. doi: 10.1371/journal.pone.0127401
- Eissenstat, D. M. (1992). Costs and benefits of constructing roots of small diameter. *J. Plant Nutr.* 15, 763–782. doi: 10.1080/01904169209364361
- Farooq, M., Gogoi, N., Barthakur, S., Baroowa, B., Bharadwaj, N., Alghamdi, S. S., et al. (2016). Drought stress in grain legumes during reproduction and grain filling. *J. Agron. Crop Sci.* doi: 10.1111/jac.12169. [Epub ahead of print].
- Girdthai, T., Jogloy, S., Kesmala, T., Vorasoot, N., Akkasaeng, C., Wongkaew, S., et al. (2009). Relationship between root characteristics of peanut in hydroponics and pot studies. *Crop Sci.* 50, 159–167. doi: 10.2135/cropsci2008.09.0529
- Hall, A. E. (2004). “Comparative ecophysiology of cowpea, common bean and peanut,” in *Physiology and Biotechnology Integration for Plant Breeding*, eds A. Blum and H. T. Nguyen (New York, NY: Marcel Dekker Inc.), 271–325.
- Huang, B., and Fry, J. (1998). Root anatomical, physiological, and morphological responses to drought stress for tall fescue cultivars. *Crop Sci.* 38, 1017–1022.
- Klaedtke, S. M., Cajiao, C., Grajales, M., Polania, J., Borrero, G., Guerrero, A., et al. (2012). Photosynthate remobilization capacity from drought-adapted common bean (*Phaseolus vulgaris* L.) lines can improve yield potential of interspecific populations within the secondary gene pool. *J. Plant Breed. Crop Sci.* 4, 49–61. doi: 10.5897/JPBSC11.087
- Lopes, M. S., and Reynolds, M. P. (2010). Partitioning of assimilates to deeper roots is associated with cooler canopies and increased yield under drought in wheat. *Funct. Plant Biol.* 37, 147–156. doi: 10.1071/FP09121
- Lord, J. M., and Westoby, M. (2012). Accessory costs of seed production and the evolution of angiosperms. *Evolution* 66, 200–210. doi: 10.1111/j.1558-5646.2011.01425.x
- Lynch, J. P. (2013). Steep, cheap and deep: an ideotype to optimize water and N acquisition by maize root systems. *Ann. Bot.* 112, 347–357. doi: 10.1093/aob/mcs293
- Lynch, J. P., Ho, M. D. (2005). Rhizoeconomics: carbon costs of phosphorus acquisition. *Plant Soil* 269, 45–56. doi: 10.1007/s11104-004-1096-4
- Mir, R. R., Zaman-Allah, M., Sreenivasulu, N., Trethowan, R. M., and Varshney, R. K. (2012). Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. *Theor. Appl. Genet.* 125, 625–645. doi: 10.1007/s00122-012-1904-9
- Muñoz-Perea, C. G., Allen, R. G., Westermann, D. T., Wright, J. L., and Singh, S. P. (2007). Water use efficiency among dry bean landraces and cultivars in drought-stressed and non-stressed environments. *Euphytica* 155, 393–402. doi: 10.1007/s10681-006-9340-z
- Passioura, J. B. (1997). “Drought and drought tolerance,” in *Drought Tolerance in Higher Plants: Genetical, Physiological, and Molecular Biological Analysis*, ed E. Belhassen (Springer Netherlands), 1–5.
- Polania, J. A., Poschenrieder, C., Beebe, S., and Rao, I. M. (2016a). Effective use of water and increased dry matter partitioned to grain contribute to yield of common bean improved for drought resistance. *Front. Plant Sci.* 7:660. doi: 10.3389/fpls.2016.00660
- Polania, J., Poschenrieder, C., Beebe, S., and Rao, I. M. (2016b). Estimation of phenotypic variability in symbiotic nitrogen fixation ability of common bean under drought stress using ¹⁵N natural abundance in grain. *Eur. J. Agron.* 79, 66–73. doi: 10.1016/j.eja.2016.05.014
- Polania, J., Rao, I. M., Beebe, S., and Garcia, R. (2009). Root development and distribution under drought stress in common bean (*Phaseolus vulgaris* L.) in a soil tube system. *Agron. Columbiana* 27, 25–32.
- Polania, J., Rao, I. M., Cajiao, C., Rivera, M., Raatz, B., and Beebe, S. (2016c). Physiological traits associated with drought resistance in Andean and Mesoamerican genotypes of common bean (*Phaseolus vulgaris* L.). *Euphytica* 210, 17–29. doi: 10.1007/s10681-016-1691-5
- Polania, J., Rao, I. M., Mejía, S., Beebe, S., and Cajiao, C. (2012). Características morfo-fisiológicas de frijol común (*Phaseolus vulgaris* L.) relacionadas con la adaptación a sequía. *Acta Agron.* 61, 197–206.

- Ramirez-Vallejo, P., and Kelly, J. D. (1998). Traits related to drought resistance in common bean. *Euphytica* 99, 127–136. doi: 10.1023/A:1018353200015
- Rao, I. M. (2014). “Advances in improving adaptation of common bean and Brachiaria forage grasses to abiotic stresses in the tropics,” in *Handbook of Plant and Crop Physiology*, ed M. Pessarakli (Boca Raton, FL: CRC Press; Taylor and Francis Group), 847–889.
- Rao, I. M., Beebe, S. E., Polania, J., Grajales, M., Cajiao, C., Ricaurte, J., et al. (2016a). Evidence for genotypic differences among elite lines of common bean in their ability to remobilize photosynthate to increase yield under drought. *J. Agric. Sci.* doi: 10.1017/S0021859616000915. [Epub ahead of print].
- Rao, I. M., Beebe, S., Polania, J., Ricaurte, J., Cajiao, C., Garcia, R., et al. (2013). Can tepary bean be a model for improvement of drought resistance in common bean? *Afr. Crop Sci. J.* 21, 265–281.
- Rao, I. M., Miles, J. W., Beebe, S. E., and Horst, W. J. (2016b). Root adaptations to soils with low fertility and aluminium toxicity. *Ann. Bot.* 118, 593–605. doi: 10.1093/aob/mcw073
- Reynolds, M. P., and Trethowan, R. M. (2007). “Physiological interventions in breeding for adaptation to abiotic stress,” in *Scale and Complexity in Plant Systems Research, Gene-Plant-Crop Relations*, eds J. H. J. Spiertz, P. C. Struik, and H. H. Van Laar (Heidelberg: Wageningen UR Frontis Series), 127–144.
- Rippke, U., Ramirez-Villegas, J., Jarvis, A., Vermeulen, S. J., Parker, L., Mer, F., et al. (2016). Timescales of transformational climate change adaptation in sub-Saharan African agriculture. *Nat. Clim. Chang.* 6, 605–609. doi: 10.1038/nclimate2947
- Rosales, M. A., Ocampo, E., Rodríguez-Valentín, R., Olvera-Carrillo, Y., Acosta-Gallegos, J. A., and Covarrubias, A. A. (2012). Physiological analysis of common bean (*Phaseolus vulgaris* L.) cultivars uncovers characteristics related to terminal drought resistance. *Plant Physiol. Biochem.* 56, 24–34. doi: 10.1016/j.plaphy.2012.04.007
- Rosales-Serna, R., Kohashi-Shibata, J., Acosta-Gallegos, J. A., Trejo-López, C., Ortiz-Cereceres, J., and Kelly, J. D. (2004). Biomass distribution, maturity acceleration and yield in drought-stressed common bean cultivars. *Field Crop. Res.* 85, 203–211. doi: 10.1016/S0378-4290(03)00161-8
- Rosas, J. C., Hernández, J. C., and Araya, R. (2003). Registration of “Bribri” small red bean (race Mesoamerica). *Crop Sci.* 43, 430.
- Ryser, P. (2006). The mysterious root length. *Plant Soil* 286, 1–6. doi: 10.1007/s11104-006-9096-1
- SAS Institute Inc (2008). SAS/STAT® 9.2.
- Sinclair, T. R. (2012). Is transpiration efficiency a viable plant trait in breeding for crop improvement? *Funct. Plant Biol.* 39, 359–365. doi: 10.1071/FP11198
- Singh, S. P., Terán, H., and Gutiérrez, J. A. (2001). Registration of SEA 5 and SEA 13 drought tolerant dry bean germplasm. *Crop Sci.* 41, 276–277. doi: 10.2135/cropsci2001.411276x
- Sponchiado, B. N., White, J. W., Castillo, J. A., and Jones, P. G. (1989). Root growth of four common bean cultivars in relation to drought tolerance in environments with contrasting soil types. *Exp. Agric.* 25, 249. doi: 10.1017/S0014479700016756
- Turner, N. C. (1979). “Drought resistance and adaptation to water deficits in crop plants,” in *Stress Physiology in Crop Plants*, eds H. Mussel and R. C. Staple (New York, NY: Wiley), 344–372.
- White, J. W. (1993). “Implications of carbon isotope discrimination studies for breeding common bean under water deficits,” in *Stable Isotopes and Plant Carbon-Water Relations*, eds E. Hall, Jr., and G. D. Farquhar (San Diego, CA: Academic press), 387–398.
- White, J. W., and Castillo, J. A. (1992). Evaluation of diverse shoot genotypes on selected root genotypes of common bean under soil water deficits. *Crop Sci.* 32, 762–765.
- White, J. W., Castillo, J. A., and Ehleringer, J. R. (1990). Associations between productivity, root growth and carbon isotope discrimination in *Phaseolus vulgaris* under water deficit. *Aust. J. Plant Physiol.* 17, 189–198.
- White, J. W., and Singh, S. P. (1991). Sources and inheritance of earliness in tropically adapted indeterminate common bean. *Euphytica* 55, 15–19. doi: 10.1007/BF00022554
- Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011). Chickpea genotypes contrasting for seed yield under terminal drought stress in the field differ for traits related to the control of water use. *Funct. Plant Biol.* 38, 270–281. doi: 10.1071/FP10244

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Polania, Rao, Cajiao, Grajales, Rivera, Velasquez, Raatz and Beebe. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of Quantitative Trait Loci Controlling Root and Shoot Traits Associated with Drought Tolerance in a Lentil (*Lens culinaris* Medik.) Recombinant Inbred Line Population

Omar Idrissi^{1,2*}, Sripada M. Udupa³, Ellen De Keyser⁴, Rebecca J. McGee⁵, Clarice J. Coyne⁶, Gopesh C. Saha⁷, Fred J. Muehlbauer⁶, Patrick Van Damme^{1,8} and Jan De Riek⁴

¹ Department of Plant Production, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium, ² Institut National de la Recherche Agronomique du Maroc (INRA), Centre Régional de Settat, Settat, Morocco, ³ International Center for Agricultural Research in the Dry Areas, Institut National de la Recherche Agronomique Morocco Cooperative Research Project, Rabat, Morocco, ⁴ Plant Sciences Unit, Applied Genetics and Breeding, The Institute for Agricultural and Fisheries Research (ILVO), Melle, Belgium, ⁵ United States Department of Agriculture, Agricultural Research Service Grain Legume Genetics and Physiology Research, Pullman, WA, USA, ⁶ United States Department of Agriculture, Agricultural Research Service Western Regional Plant Introduction, Washington State University, Pullman, WA, USA, ⁷ Brotherton Seed Company, Washington, DC, USA, ⁸ Faculty of Tropical AgriSciences, Czech University of Life Sciences, Prague, Czech Republic

OPEN ACCESS

Edited by:

Maria Carlota Vaz Patto,
Universidade Nova de Lisboa,
Portugal

Reviewed by:

Maoteng Li,
Huazhong University of Science and
Technology, China
Rebeca Iglesias-Garcia,
Nebrija University, Spain

*Correspondence:

Omar Idrissi
Omar.Idrissi@UGent.be;
o.idrissi@yahoo.fr

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 31 March 2016

Accepted: 21 July 2016

Published: 23 August 2016

Citation:

Idrissi O, Udupa SM, De Keyser E, McGee RJ, Coyne CJ, Saha GC, Muehlbauer FJ, Van Damme P and De Riek J (2016) Identification of Quantitative Trait Loci Controlling Root and Shoot Traits Associated with Drought Tolerance in a Lentil (*Lens culinaris* Medik.) Recombinant Inbred Line Population. *Front. Plant Sci.* 7:1174. doi: 10.3389/fpls.2016.01174

Drought is one of the major abiotic stresses limiting lentil productivity in rainfed production systems. Specific rooting patterns can be associated with drought avoidance mechanisms that can be used in lentil breeding programs. In all, 252 co-dominant and dominant markers were used for Quantitative Trait Loci (QTL) analysis on 132 lentil recombinant inbred lines based on greenhouse experiments for root and shoot traits during two seasons under progressive drought-stressed conditions. Eighteen QTLs controlling a total of 14 root and shoot traits were identified. A QTL-hotspot genomic region related to a number of root and shoot characteristics associated with drought tolerance such as dry root biomass, root surface area, lateral root number, dry shoot biomass and shoot length was identified. Interestingly, a QTL (QRSratio_{IX-2.30}) related to root-shoot ratio, an important trait for drought avoidance, explaining the highest phenotypic variance of 27.6 and 28.9% for the two consecutive seasons, respectively, was detected. This QTL was closed to the co-dominant SNP marker TP6337 and also flanked by the two SNP TP518 and TP1280. An important QTL (QLRN_{III-98.64}) related to lateral root number was found close to TP3371 and flanked by TP5093 and TP6072 SNP markers. Also, a QTL (QSRL_{IV-61.63}) associated with specific root length was identified close to TP1873 and flanked by F7XEM6b SRAP marker and TP1035 SNP marker. These two QTLs were detected in both seasons. Our results could be used for marker-assisted selection in lentil breeding programs targeting root and shoot characteristics conferring drought avoidance as an efficient alternative to slow and labor-intensive conventional breeding methods.

Keywords: lentil, ecophysiology, drought tolerance, breeding, plant, QTL, marker-assisted selection

INTRODUCTION

Lentil (*Lens culinaris* Medik.) is an important grain legume crop that is often grown in sustainable farming systems and for nutrition in the world. Its ability to enhance soil fertility through atmospheric nitrogen fixation allows substantial reduction in fertilizer use and significant production improvement in cereal-based cropping systems thanks to the benefits of rotation.

Lentil grains are a rich source of proteins and some important micronutrients such as iron and zinc (Grusak and Coyne, 2009; Thavarajah et al., 2011). Consumed as staple food in developing countries and as vegetarian dishes elsewhere, lentil grains are considered a very healthy food. The United Nations, in its 68th General Assembly, declared 2016 as the International Year of Pulses (annual leguminous crops harvested for dry grains) in order to highlight the nutritional benefits of pulses as part of sustainable food production aimed towards food and nutrition security (FAO, 2015).

In the arid and semi-arid areas and also in the context of climate change and global warming, drought is one of the major constraints that can limit lentil production and cause substantial yield losses (Malhotra et al., 2004; Stoddard et al., 2006; Sarker et al., 2009). Developing cultivars with enhanced drought tolerance by conventional breeding often has limited success due to the complexity of this trait and the difficulties with finding reliable and suitable phenotyping methods. For example, well-developed root systems have been shown to be linked to drought tolerance as an avoidance mechanism guaranteeing plant productivity under water-limited conditions (Kashiwagi et al., 2005; Sarker et al., 2005; Verslues et al., 2006; Gaur et al., 2008; Vadez et al., 2008; Aswaf and Blair, 2012; Comas et al., 2013; Idrissi et al., 2015a,b). However, it is difficult to screen large numbers of accessions for these root traits using conventional methods. Thus, applying a marker-assisted selection for these traits would offer an interesting alternative in breeding programs targeting drought tolerance. As such, identifying and mapping DNA markers linked to genes controlling rooting patterns associated with drought tolerance will assist in reliable and efficient identification and development of tolerant cultivars. Several studies have shown that root traits are polygenically controlled, whereas they also identified related quantitative trait loci (QTLs) for different species such as maize (Ruta, 2008), common bean (Cichy et al., 2009; Aswaf and Blair, 2012), barley (Sayed, 2011), soybean (Brensha et al., 2012) and chickpea (Kashiwagi et al., 2014).

Lentil has a genome size of about 4 Gbp (Arumuganathan and Earle, 1991); several kinds of DNA markers have been developed and mapped, including RAPDs, ISSRs, AFLPs, SRAPs, SSRs, and SNPs (Eujayl et al., 1998; Rubeena et al., 2003; Hamwieh et al., 2005; Saha et al., 2010; Sharpe et al., 2013). Idrissi et al. (2015a) confirmed evidence of high genetic variability, high heritability and polygenic control of root and shoot characteristics. However, to our knowledge, no QTLs related to root traits have been reported for lentil to date. Thus, the objective of this study was to identify and map QTLs related to root and shoot traits associated with drought tolerance in a lentil recombinant inbred

line population (RIL) as a promising step towards a marker-assisted selection approach. It also aimed to investigate the stability of detected QTLs by performing the analysis on two consecutive seasons.

MATERIALS AND METHODS

Plant Materials

A recombinant inbred line (RIL) population developed from a cross between two contrasting parents, ILL6002 and ILL5888 (Saha et al., 2010), obtained from Fred J. Muehlbauer, USDA-ARS, Washington State University, Pullman, USA, was used in this study. The RIL population consisted of the two parents and 132 F₆₋₈ lines. The lines were advanced to the F₆₋₈ generation from individual F₂ using single seed descent. The ILL6002 parent is a vigorous line reported as drought tolerant and with a well-developed root system (Sarker et al., 2005; Singh et al., 2013). On the other hand, ILL5888 is a drought sensitive line and has a less-developed root system and vegetative biomass. The two parents also differ in disease resistance (*Stemphylium* blight), flowering and maturity time, seed diameter, 100-seed weight, growth habit and plant height (Saha et al., 2010, 2013).

RIL Root and Shoot Traits Phenotyping and Drought Tolerance Evaluation

This F₆₋₈ population was previously characterized for root and shoot traits related to drought tolerance (Idrissi et al., 2015a). Briefly, the population was evaluated under greenhouse conditions for root and shoot traits associated with drought tolerance under two contrasting watering regimes (well-watered and progressive drought-stressed) using the standard nutrition solution EEG MESTSTOF 19-8-16 (4) for two consecutive growing seasons (2013 and 2014). A completely randomized block design with three replications was used. Four uniformly germinated seeds were planted in plastic pots (H 35 × D 24 cm) filled with fine perlite (diameter ≤ 2 mm) in order to be able to extract intact roots without damage (Day, 1991; Rabah Nasser, 2009). The initial moisture in all the pots of both watering regimes was 75% of field capacity. It decreased to about 22% for the drought-stressed regime where plants were watered only once in the beginning of the experiment, while it was maintained at 75% for the well-watered treatment by watering plants twice a week as described in Idrissi et al. (2015a). At 38 days after sowing, plants were carefully extracted without damage to the roots, then shoots and roots were separated into plastic bags. Washed roots were preserved in a refrigerator (4°C, 90% relative humidity) to avoid drying before being scanned using an EPSON Scan scanner. The images were then analyzed using Image J software (Abramoff et al., 2004) combined with Smart Roots software (Lobet et al., 2011). From the scanned images, taproot length (TRL; cm plant⁻¹), average taproot diameter (TRD; mm plant⁻¹), root surface area (RSA; cm² plant⁻¹) and lateral root number (LRN) were measured. Dry root and shoot biomass (DRW, DSW; mg plant⁻¹) were measured after oven-drying at 72°C for 48 h. Chlorophyll content was estimated according to the SPAD values measured at 32 days after sowing using a SPAD-502Plus chlorophyll meter (Konica Minolta, Japan), four

measures were taken in fully expanded leaves per plant. The wilting score (WS) corresponding to the degree of wilting severity was used to estimate drought tolerance using the following 0–4 score scale (Singh et al., 2013): 0 = healthy plants with no visible symptoms of drought stress; 1 = green plants with slight wilting; 2 = leaves turning yellowish green with moderate wilting; 3 = leaves yellow–brown with severe wilting; and 4 = completely dried leaves and/or stems. Seedling vigor (SV) was recorded following the 1–5 IBPGR and ICARDA (1985) scale: 1 = very poor; 2 = poor; 3 = average; 4 = good; 5 = excellent. Root–shoot ratio (RS ratio) was calculated by dividing the dry root weight by the dry shoot weight. Growth rate (GR; cm) was estimated as the gain of length between 12 (SL12DAS; cm) and 22 days after sowing (SL22DAS; cm; GR = SL22DAS–SL12DAS). Specific root length (SRL) and specific root surface area (SRSA) were estimated by dividing root length and root surface area, respectively, by dry root weight. All the measures were recorded as the mean value based on the four plants per individual genotype in each pot. A summary of genetic variation and heritability of these traits is provided in Table 4 (Supplementary Material).

RIL Genotyping

The previously developed linkage map of Saha et al. (2010) was created based on the same RIL population used in this study (ILL6002 × ILL5888). The initial mapping data of Saha et al. (2010) which consisted of 23 SSR, 108 SRAP, and 30 RAPD markers, were kindly provided by the authors. The map was further enhanced using 220 polymorphic Single Nucleotide Polymorphism (SNP) markers developed using the Genotyping By Sequencing (GBS) technique and 180 polymorphic Amplified Fragment Length Polymorphism (AFLP) markers.

Genotyping by Sequencing for SNP Identification

SNP data were obtained from 92 (out of 132) RILs using GBS. The GBS procedure of Poland et al. (2012) was used, including their 48 bar-coded adapters with a *Pst* I overhang; genomic DNA was digested with the enzymes *Pst* I and *Msp* I. The ligation reaction was completed using bar-coded Adapter 1 and the common Y-adapter in a master mix of buffer, ATP and *T*₄-ligase. Ligated samples were pooled and PCR-amplified in a single tube, producing libraries of 48 samples each. The libraries were sequenced on two lanes of Illumina HiSeq2000 (University of California Berkeley V.C. Genomic Sequencing Lab). The sequencing data were processed to remove low quality data using in-house scripts and analyzed using Stacks software (Catchen et al., 2011, 2013). Two hundred twenty SNPs that proved to be polymorphic between both parents of the RIL population ILL6002 × ILL5888 (Wong et al., 2015) were analyzed.

AFLP Genotyping

The AFLP protocol of Vos et al. (1995) with minor modifications (De Riek et al., 2001) was performed as described in Idrissi et al. (2015c). Out of 12 primer combinations tested, seven (*Eco*RI-ACA + *Mse*I-CAG, *Eco*RI-ACA + *Mse*I-CTG, *Eco*RI-ACA

+ *Mse*I-CTT, *Eco*RI-ACG + *Mse*I-CAA, *Eco*RI-AGC + *Mse*I-CAA, *Eco*RI-AGC + *Mse*I-CAG, *Eco*RI-AGC + *Mse*I-CTG) were selected and used for genotyping the RIL population.

Linkage Analysis and Map Construction

Five hundred sixty-one molecular markers on 132 RILs were used for linkage analysis and construction of a linkage map using JoinMap[®] 4 (Van Ooijen, 2006; Table 1). First, segregation according to Mendelian expectation ratio of 1:1 was tested using the chi-square test at a significance level of 0.05, markers with distorted segregation were removed prior to further analysis. The grouping tree of the JoinMap[®] program was calculated using independent LOD (Logarithm of odds) as grouping parameter with threshold ranges of 6 for start and 30 for end, and 1 for step. Stable sets of markers at higher LOD values were selected. After initial creation of groups, the *Strongest Cross Link* (SCL) information from the output results was used for inspecting assignment of markers to groups, those with SCL-values larger than 5, indicating that they have strong linkage outside their respective groups, were assigned to the corresponding groups. This was repeated until all markers of each group had SCL-values smaller than 5. Linkage groups were calculated using the maximum likelihood mapping algorithm with default values as in the software. Map order in each linkage group was verified using the regression mapping algorithm with the following parameters: LOD threshold larger than 4, recombination frequency smaller than 0.25, Kosambi function as mapping function for genetic distance calculation and the second round map of the algorithm. The final linkage map was generated using MapChart[®] 2.3 program (Voorrips, 2002).

QTL Analysis

QTL analysis was performed for each season separately for drought-stressed treatments in order to check the stability of detected QTLs using MapQTL[®] 5 program (Van Ooijen, 2004). First, Kruskal-Wallis test was performed to determine a set of markers linked to each quantitative trait. Simple Interval Mapping was performed to identify linkage groups and positions with significant LOD scores. For each trait, LOD score threshold was determined based on a permutation test using 1000 iterations at a *P*-value of 0.05; LOD scores above these values were considered as significant. Co-factor selection was performed based on automatic co-factor selection implemented in the software for each linkage group and on manual selection of individual markers with significant LOD scores from Simple

TABLE 1 | Marker types used for linkage map development.

Marker types	Number of polymorphic markers	Final number of mapped markers
SNP	220	106
SSR	23	13
SRAP	108	56
AFLP	180	60
RAPD	30	17

Interval Mapping output before applying Multiple-QTL Models (MQM) mapping (also called Composite Interval Mapping). Performing MQM mapping with markers close to significant LOD score positions as co-factors allows reduction of residual variance, thus enhancing the power of QTL detection. For each quantitative trait, co-factor selection and MQM mapping were repeated until no further enhancement was obtained (no more QTLs detected, increase in LOD scores and explained variances). From the MQM mapping output, closest marker, flanking markers, additive effect and percentage of explained variance for each detected QTL and for each quantitative trait were determined for seasons. Final results, with significant LOD scores and intervals, for each detected QTL per linkage group, were generated using MapChart[®] 2.3 program (Voorrips, 2002).

All detected QTLs were named as follows: Q'Trait name abbreviation', "linkage group number" – "position in cM". For example: QLRN_{III-98.64} is a QTL associated with LRN identified in linkage group III at position 98.64 cM.

RESULTS

GBS for SNP Identification

Selection of the genotyping-by-sequencing two enzyme method of Poland et al. (2012) and the enzymes *Msp I* and *Pst I* was based on the results of Wong et al. (2015) lentil SNP discovery across the lentil species. Using GBS, 220 polymorphic SNPs were deemed high quality for mapping, after satisfying quality control filtering based on deleting low quality and redundant SNPs using haplotype information for read depth (3), lack of redundancy and segregation in the parents. Genome coverage was reasonable, but incomplete, across six linkage groups (LG I, LG II, LG III, LG IV, LG VI, and LG IX; Figure 1).

Linkage Analysis and Map Construction

Marker distortion tested by Chi-square test ($P < 0.05$) revealed that 35.4% of SNPs, 43% of SSRs, 18% of SRAPs, 52.7% of AFLPs and 20% of RAPDs did not segregate according to the expected 1:1 ratio and were removed from the analysis. Out of 17 stable groups selected from the grouping tree, a total of 252 out of the 561 polymorphic markers were finally mapped in nine linkage groups spanning a total length of 2022.8 cM (Tables 1, 2). Final linkage groups were established using the SCL information. Linkage group length ranged from 71.7 to 531.1 cM whereas average distance between two markers ranged from 5.12 (LG V) to 9.8 cM (LG II) (Table 2; Figure 1). Seven linkage groups had a length of more than 100 cM (LG I, LG II, LG III, LG IV, LG VI, LG VII, and LG IX). Both co-dominant (SNP and SSR) and dominant (SRAP, AFLP, and RAPD) markers were present in six linkage groups, while three linkage groups (LG V, LG VII, and LG VIII) were composed only out of dominant markers.

QTL Identification

A total number of 18 QTLs associated with 14 root and shoot traits were detected under drought-stressed

conditions during two seasons (Table 3; Figures 2, 3). LOD score, percentage of explained phenotypic variance and additive effect of detected QTLs ranged from 2.75 (TRL) to 8.14 (DSW), from 4.3 (QRSrati_{IX-77.72}) to 28.9% (QRSrati_{IX-2.30}) and from -5.17 (LRN) to 8.10 (DRW), respectively.

Seven of the detected QTLs were co-located on LG VII at position 21–22 cM, with UBC34 as the closest marker and ME5XR10–UBC1 as the two flanking markers: QDRW_{VII-21.94}, QLRN_{VII-21.94}, QRSA_{VII-21.94}, QDSW_{VII-21.94}, QSL12_{VII-20.75}, QSL22_{VII-21.75}, and QGR_{VII-21.94}.

Among the 18 detected QTLs, 12 were evidenced for the drought-stressed treatment for both seasons: QDRW_{VII-21.93}, QRSA_{VII-21.94}, QDSW_{VII-22.94}, QRSrati_{IX-2.30}, QSL12_{IV-103.83}, QSL12_{VI-170.87}, QSL12_{VII-19.71}, QSL22_{VII-21.94}, QLRN_{III-98.64}, QLRN_{VII-21.94}, QSRL_{IV-61.63}, and QSPAD_{VIII-72.15}. Interestingly, among these stable QTLs, QRSrati_{IX-2.30}, located at 2.30 cM on LG IX, is associated with a high root-shoot ratio and had LOD scores of 6.20 and 5.11 for 2013 and 2014 seasons, respectively. The explained phenotypic variance of this QTL was the highest with 27.6 and 28.9% and an additive effect of 1.23 and 1.84 for 2103 and 2014 seasons, respectively. The closest marker to this QTL is SNP marker TP6337 located at 2.3 cM whereby the two flanking markers are TP518 and TP1280, located respectively at 0 and 2.9 cM.

Two QTLs were identified for dry root biomass, QDRW_{VII-21.93}, accounted for 22.2% (with a LOD score of 7.21) and 21.3% (with a LOD score of 6.88) of the phenotypic variance with additive effects of 8.10 and 7.47 for 2013 and 2014 seasons, respectively.

Among the three QTLs detected for LRN, QLRN_{III-98.64}, was located at 98.64 cM position on LG III close to TP3371 SNP marker and flanked by the two SNP markers TP5093–TP6072. The LOD scores, percentage of explained phenotypic variances and additive effects were 2.94, 23.5% and -5.17, and 3.31, 24%, and -5.15 for 2013 and 2014 seasons, respectively. An important QTL was also identified for SRL, namely QSRL_{IV-61.63}, that was detected for both seasons with LOD scores, percentage of explained phenotypic variances and additive effects of 3.84, 16.8% and 0.83 and 3.63, 16.2% and 0.32, respectively, for 2013 and 2014.

Three QTLs were identified to be linked to chlorophyll content in which one was common for both seasons. The latter is the QTL QSPAD_{VIII-72.15}, which was detected with LOD scores, percentage of explained phenotypic variances and additive effects of respectively 3.98, 10.7% and -2.20 for 2013 season and 4.25, 13.1% and -2.20 for 2014 season.

Also, a QTL related to early vegetative vigor estimated by SV was detected for the 2013 experiment. This QTL, QSV_{VII4}, was located on LG VII at position 4 cM, had a LOD score of 3.46, an additive effect of 0.29 and explained 14.9% of total phenotypic variance.

A QTL QWS_{I-22.53}, related to drought tolerance as estimated by WS, is located at 22.53 cM position on LG I with a LOD score of 3.08 and 18.8% as percentage of explained phenotypic variance.

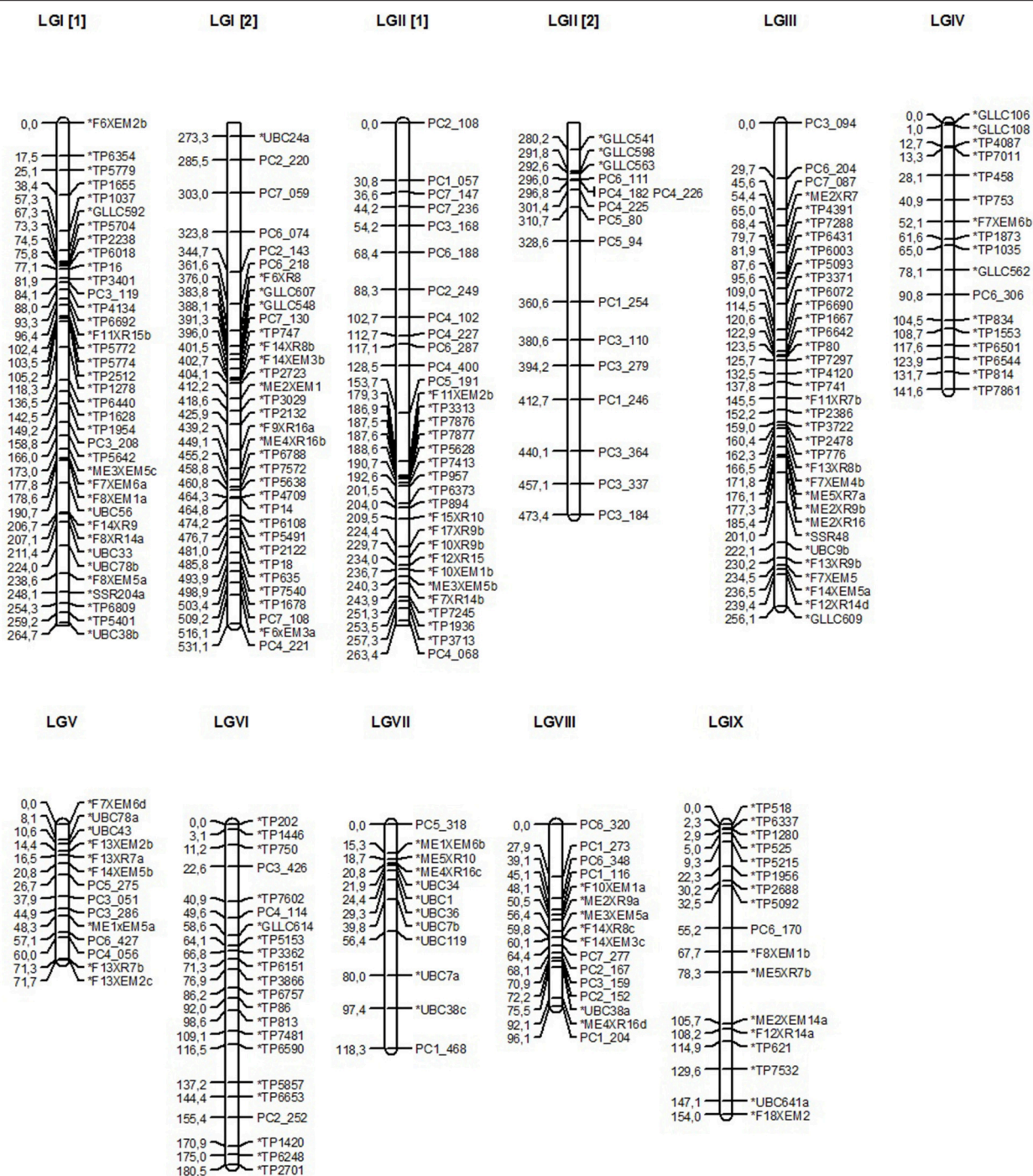


FIGURE 1 | Genetic linkage map of lentil developed at LOD score > 6 using maximum likelihood mapping algorithm of JoinMap® 4 program and drawn using MapChart® program. LGI-LGIX correspond to linkage groups, marker names are presented right to the linkage group and the genetic positions in CentiMorgans (cM) left. SNP markers are denoted by *TP_, SSR by *GLLC_, SRAP starting by *M_ or F_, RAPD by *UBC_ and AFLP by PC_.

DISCUSSION

The genetic linkage map of lentil initially developed by Saha et al. (2010) using a ILL6002 × ILL5888 RIL population containing 139 markers and 14 linkage groups was enhanced by adding

SNP co-dominant markers and AFLP dominant markers, thereby increasing marker density and total spanned length. The number of linkage groups was reduced to nine with a total number of 252 mapped markers covering 2022.8 cM compared to 1565.2 cM in the previous genetic map. Average distance between markers

TABLE 2 | Linkage groups of the developed lentil linkage map and marker distribution.

Linkage groups	Number of mapped markers	Length (cM)	Average distance between markers (cM)
LGI	71	531.1	7.4
LGII	48	473.4	9.8
LGIII	35	256.1	7.3
LGIV	17	141.6	8.3
LGV	14	71.7	5.12
LGVI	22	180.5	8.2
LGVII	12	118.3	9.8
LGVIII	16	96.1	6.0
LGIX	17	154.0	9.0
Total	252	2022.8	8.0

was reduced from 11.3 to 8 cM. Sharpe et al. (2013) reported a lentil map with seven linkage groups using SNP and SSR markers. 62.77% of markers from the Saha et al. (2010) linkage map were also mapped in the genetic map developed in our study. Several sets of markers from the previous genetic map were confirmed to be linked to each other in our map. For instance, all markers from LG 1 from the map of Saha et al. (2010) were also mapped in LG I of our map. Thirteen markers out of a total of 19 mapped in LG 2 were mapped in LG III and four in LG IV of our map. Nine markers from LG 3 were mapped in LG V of our map and all those from LG 4 except for two that ended up in LG VII of our enhanced map. All markers from LG 11 of the previous map (except two) were mapped in LG II. All markers of LG 13 and LG 14 were mapped in LG IX and LG VIII of our map, respectively. Our linkage groups could not be assigned per Sharpe et al. (2013), the best lentil linkage map with seven linkage groups likely corresponding to the seven chromosomes of the genome developed to date, due to lack of common markers. We used a combination of dominant and co-dominant markers to develop a linkage map with reduced gaps. Since SNP data were not available for the whole population, we added also dominant AFLP markers for map construction. In other studies, dominant markers were also used together with co-dominant ones for the development of linkage maps and QTL analysis to overcome different limits such as genetic marker availability and large gaps in linkage groups (Gaudet et al., 2007; De Keyser et al., 2010; Kaur et al., 2014; Muys et al., 2014; Ting et al., 2014). Although, maximum likelihood mapping algorithm often results in increased map length, it is considered to be more robust with missing data, genotyping errors and the use of markers with low information content (Lincoln and Lander, 1992; Van Ooijen, 2006; Cartwright et al., 2007; De Keyser et al., 2010). This algorithm uses multipoint analysis to approximate missing genotypes using nearby markers (De Keyser et al., 2010). Genetic linkage maps based on this approach giving the most likely marker order (De Keyser et al., 2010) were reported to be suitable for QTL mapping (Kim, 2007; De Keyser et al., 2010). Thus, we adopted this approach as the main objective of our study was to identify QTLs related to root and shoot traits. Furthermore, although we used dominant markers such as AFLPs known to result in longer map, our

linkage groups did not have extreme lengths and the total map length of 2022.8 cM is among common reported values in similar studies on lentil. Duran and Perez De La Vega (2004) reported a genetic linkage map of 2172 cM length using SSR, AFLP, ISSR and RAPD markers. Gupta et al. (2012) used SSR, ISSR and RAPD markers to construct a map of 3843.4 cM length. Also, Kaur et al. (2014) used SSR and SNP markers to develop a map of 1178 cM length. Using predominantly SNP markers and few SSRs, Sharpe et al. (2013) constructed a shorter map of 834.7 cM length. More recently, Ates et al. (2016) developed a map spanning a total length of 4060.6 cM and composed of seven linkage groups using SSR and SNP markers to identify QTLs controlling genes for Selenium uptake in lentil.

High genetic variability, quantitative, continuous and normally distributed variation as well as high heritability estimate values of all studied traits were reported in Idrissi et al. (2015a).

In all, 18 QTLs were identified for root and shoot traits for both seasons under progressive drought-stressed treatments in the lentil RIL population ILL6002 × ILL5888. Among these QTLs, 12 were evidenced for both seasons. Aswaf and Blair (2012) reported a total of 15 putative QTLs for seven rooting pattern traits and four shoot traits under drought-stressed treatments in common bean (*Phaseolus vulgaris* L.). Varshney et al. (2014) reported drought tolerance-related root trait QTLs in chickpea (*Cicer arietinum* L.). In soybean (*Glycine max* L.), Manavalan et al. (2015) identified a QTL region controlling a number of root and shoot architectural traits. In lentil, to our knowledge, this is the first report on QTLs related to root and shoot traits associated with drought tolerance. Interestingly, QTL QRSratio_{IX-2.30} related to root-shoot ratio, an important trait for drought avoidance (Verslues et al., 2006), was confirmed to be present on LG IX at 2.30 cM position during the two seasons. Among detected QTLs, this QTL explained the highest percentage of phenotypic variance and was close to the co-dominant SNP marker TP6337 (C/T) and furthermore was flanked by the two SNP markers TP518 (A/G) and TP1280 (G/T). These markers are potentially important for their practical use for marker-assisted selection in breeding programs targeting drought tolerance. It should be pointed out that the same SNP markers were confirmed as being linked to root-shoot ratio when using only SNP markers on 92 RILs for linkage map construction and QTL mapping (data not shown).

A QTL-“hotspot” genomic region was identified on LG VII close to UBC34 RAPD marker and ME4XR16c SRAP marker, and was identified to be linked to the genetic control of a number of root and shoot traits for both seasons: DRW, LRN, RSA, DSW, and SL at 12 and 22 days after sowing. These traits were shown to be significantly correlated (Idrissi et al., 2015a). Similarly, a QTL-“hotspot” related to 12 root traits was reported by Varshney et al. (2014) in chickpea (*Cicer arietinum* L.). Although practical efficient use of the identified genomic region in the ILL6002 × ILL5888 lentil population for marker-assisted selection could be limited by the dominant character of the closest RAPD marker, SRAP markers identified close to this genomic region could be used for assisting in the selection for linked traits. SRAP markers targeting the coding regions of open-reading frames of

TABLE 3 | Characteristics of quantitative trait loci (QTL) identified under progressive drought stress in the RIL population (ILL6002 × ILL5888) for the 2013 and 2014 seasons.

Trait*	QTL**	LOD Score ***	PVE% ****	Additive effect *****	Closest marker	Flanking markers
DROUGHT-STRESSED TREATMENT (2013)						
DRW	<u>QDRW_{VII-21.94}</u>	7.21	22.2	8.10	UBC34	ME5XR10 - UBC1
LRN	<u>QLRN_{III-98.64}</u>	2.94	23.5	-5.17	TP3371	TP5093 - TP6072
	<u>QLRN_{VII-21.94}</u>	2.91	5.3	2.78	UBC34	PC5_318 - UBC36
RSA	<u>QRSA_{VII-21.94}</u>	4.36	14.1	1.68	UBC34	ME5XR10 - UBC1
DSW	<u>QDSW_{VII-21.94}</u>	8.14	25.7	5.18	UBC34	ME5XR10 - UBC1
RS ratio	<u>QRSratio_{IX-2.30}</u>	6.20	27.6	1.23	TP6337	TP518 - TP1280
SRL	<u>QSRL_{IV-61.63}</u>	3.84	16.8	0.83	TP1873	F7XEM6b - TP1035
	QSRL _{VII-31.25}	2.83	10.2	-0.03	UBC36	ME4XR16c - UBC7b
SL12DAS	<u>QSL12_{IV-102.83}</u>	4.02	16.5	-0.54	TP834	PC6_306 - TP1553
	<u>QSL12_{VI-170.87}</u>	3.58	15.9	0.50	TP1420	PC2_252 - TP6248
	<u>QSL12_{VII-20.75}</u>	2.90	8.1	0.38	ME4XR16c	ME5XR10 - UBC34
SL22DAS	<u>QSL22_{VII-21.75}</u>	4.55	12.2	0.78	UBC34	ME5XR10 - UBC1
	QSL22 _{IV-102.83}	4.22	19.2	-0.94	TP834	PC6_306 - TP1553
GR	QGR _{VII-21.94}	2.82	9.4	0.41	UBC34	ME5XR10 - UBC1
SV	QSV _{VII-4}	3.46	14.9	0.29	PC5_318	PC5_318 - ME1XEM6b
SPAD	<u>QSPAD_{VIII-72.15}</u>	3.98	10.7	-2.20	PC2_152	PC3_159 - UBC38a
	QSPAD _{I-158.76}	3.41	9.2	2.14	PC3_208	TP1954 - TP5642
WS	QWS _{I-22.53}	3.08	18.8	0.46	TP5779	TP6354 - TP1655
DROUGHT-STRESSED TREATMENT (2014)						
DRW	<u>QDRW_{VII-21.93}</u>	6.88	21.3	7.44	UBC34	ME5XR10 - UBC1
LRN	<u>QLRN_{III-98.64}</u>	3.31	24	-5.15	TP3371	TP5093 - TP6072
	<u>QLRN_{VII-21.94}</u>	2.89	10	6.98	UBC34	ME5XR10 - UBC1
RSA	<u>QRSA_{VII-21.94}</u>	4.24	13.8	1.54	UBC34	ME5XR10 - UBC36
DSW	<u>QDSW_{VII-22.94}</u>	6.96	20.7	4.40	UBC34	ME5XR10 - UBC1
	QDSW _{IX-73.72}	2.90	9	-2.89	ME5XR7b	F8XEM1b - ME2XEM14a
RS ratio	<u>QRSratio_{IX-2.30}</u>	5.11	28.9	1.84	TP6337	TP518 - TP1280
	QRSratio _{IX-77.72}	3.45	4.3	0.14	ME5XR7b	F8XEM1b - ME2XEM14a
	QRSratio _{III-49.62}	2.95	14.7	-0.14	PC7_087	PC6_204 - ME2XR7
TRL	QTRL _{IV-52.11}	2.75	9.4	1.44	F7XEM6b	TP753 - TP1873
SRL	<u>QSRL_{IV-61.63}</u>	3.63	16.2	0.32	TP1873	F7XEM6b - TP1035
TRD	QTRD _{IX-111.15}	3.39	12.9	-0.04	F12XR14a	ME2XEM14a - TP621
SL12DAS	<u>QSL12_{VI-170.87}</u>	3.64	15.8	-0.48	TP6248	TP1420 - TP2701
	<u>QSL12_{IV-103.83}</u>	3.55	13.5	-0.45	TP834	PC6_306 - TP1553
	<u>QSL12_{VII-19.71}</u>	2.94	8.1	0.35	ME5XR10	ME1XEM6b - ME4XR16c
SL22DAS	<u>QSL22_{VII-21.94}</u>	4.28	12.1	0.65	UBC34	ME5XR10 - UBC1
SPAD	<u>QSPAD_{VIII-72.15}</u>	4.25	13.1	-2.20	PC2_152	PC3_159 - UBC38a

*DRW, dry root weight (mg plant^{-1}); LRN, lateral root number; TRL, taproot length (cm plant^{-1}); SRL, specific root length ($\text{cm mg}^{-1} \text{ plant}^{-1}$); TRD, average taproot diameter (mm plant^{-1}); RSA, root surface area ($\text{cm}^2 \text{ plant}^{-1}$); DSW, dry shoot weight (mg plant^{-1}); SL12DAS, shoot length at 12 days after sowing (cm plant^{-1}); SL22DAS, shoot length at 22 days after sowing (cm plant^{-1}); GR, growth rate (cm plant^{-1}); SV, seedling vigor; SPAD, chlorophyll content; RS ratio, root-shoot ratio; WS, wilting score.

**Bolded and underlined QTLs were identified for both seasons.

***The presence of QTL was declared when the LOD score is above the threshold value obtained by a permutation test for each quantitative trait.

****PVE: percentage of variance explained.

*****Positive values of additive effect mean that positive allele comes from the ILL6002 parent, while negative values mean that positive allele comes from the ILL5888 parent.

the genome, considered as better than RAPDs and technically less challenging than AFLPs, are of potential interest for QTL mapping (Chen et al., 2007; Yuan et al., 2008; Zhang et al., 2009; Saha et al., 2010, 2013; Robarts and Wolfe, 2014). Furthermore, up to 20% of SRAP markers were found to be co-dominant (Li and Quiros, 2001). Dry root weight, reported to be associated with drought tolerance by Idrissi et al. (2015a) in lentil, and other

root and shoot traits such as root surface area and dry shoot weight also associated with drought tolerance are linked to this “hotspot” genomic region.

QTL QLRN_{III-98.64}, related to LRN located at 98.64 cM position on LG III, was identified during both seasons explained 23.5 and 24% variations for 2013 and 2014, respectively. This QTL was close to SNP marker TP3371 (C/T) whereas its

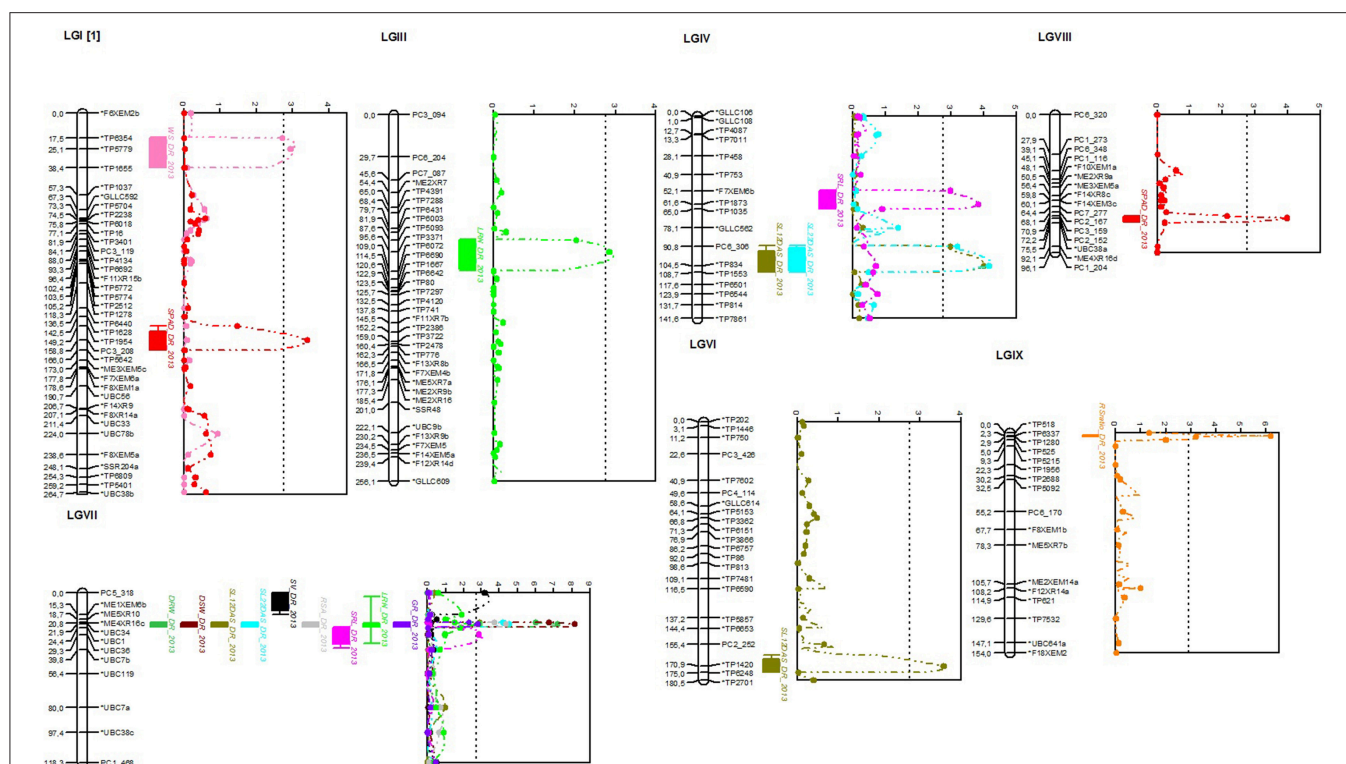
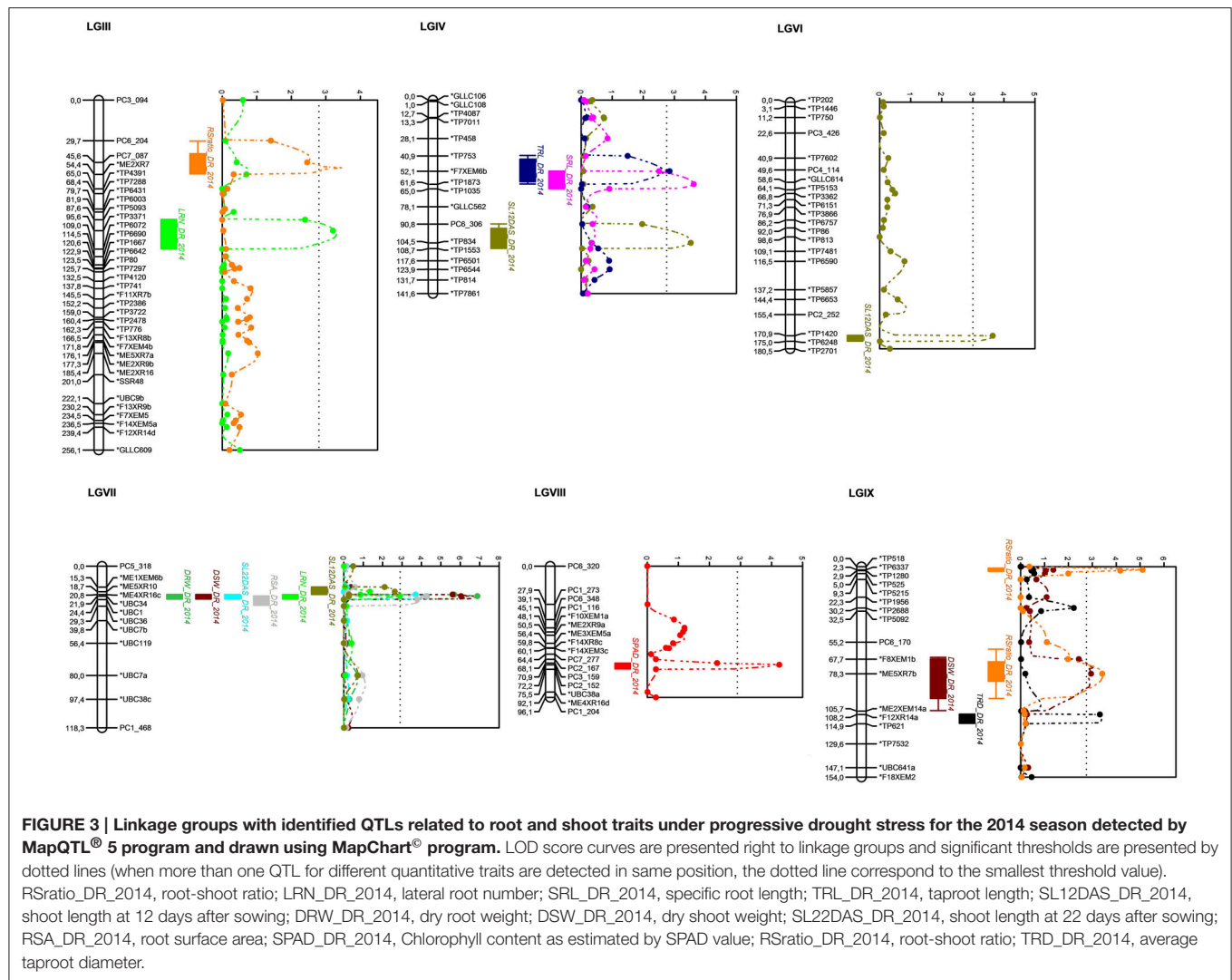


FIGURE 2 | Linkage groups with identified QTLs related to root and shoot traits under progressive drought stress for the 2013 season detected by MapQTL® 5 program and drawn using MapChart® program. LOD score curves are presented right to linkage groups and significant thresholds are presented by dotted lines (when more than one QTL for different quantitative traits are detected in same position, the dotted line correspond to the smallest threshold value). WS_DR_2013, wilting score; SPAD_DR_2013, chlorophyll content as estimated by SPAD value; LRN_DR_2013, lateral root number; SL12DAS_DR_2013, shoot length at 12 days after sowing; SL22DAS_DR_2013, shoot length at 22 days after sowing; SRL_DR_2013, specific root length; DRW_DR_2013, dry root weight; DSW_DR_2013, dry shoot weight; SV_DR_2013, seedling vigor; RSA_DR_2013, root surface area; GR_DR_2013, growth rate; RRatio_DR_2013, root-shoot ratio.

significant interval is between TP5093 (C/T) and TP6072 (A/G) SNP markers. Thus, the efficient use of these markers in breeding programs is possible for screening for higher LRN. High lateral root number was previously reported to be associated with drought tolerance and yield in lentil under drought stress (Sarker et al., 2005). Similarly, QTL QSRL_{IV-61.63}, located at 61.63 cM on LG IV and related to SRL, was detected in both seasons with fairly high LOD scores of 3.84 and 3.63 for 2013 and 2014 respectively. This QTL, explaining 16.8% of phenotypic variance, was close to TP1873 (A/C) SNP marker (61.6 cM) and flanked by the two F7XEM6b SRAP (52.1 cM) and TP1035 (A/T) SNP markers (65 cM). Specific root length is considered an important root trait that can contribute to plant productivity under drought (Comas et al., 2013). Therefore, the use of these linked markers to screen lines with longer root length should be of potential interest. Three QTLs were identified for chlorophyll content as estimated by the SPAD value. Among them, QSPAD_{I-158.76} is located at 158.76 cM on LG I close to AFLP marker PC3_208 and flanked by the two co-dominant SNP markers TP1954 (A/T) and TP5642 (A/T) that could be efficiently used in marker-assisted selection. Idrissi et al. (2015a) reported correlations of SPAD value of 0.46 and 0.48 with dry root biomass and drought tolerance, respectively, in the same mapping population used here.

A QTL QWS_{I-22.53} related to drought tolerance as estimated by the WS, located at 22.53 cM on LG I and explaining 18.8% of total phenotypic variance, is close to TP5779 (A/T) and flanked by TP6354 (C/T) and TP1655 (C/T) SNP markers, was identified during the 2013 season. After validation, these markers could be used for screening for drought tolerance. Wilting severity due to drought stress was reported to be correlated with relative water content in lentil (Idrissi et al., 2015b) indicating the importance of this parameter for the identification of drought tolerant cultivars. QTLs for drought tolerance as estimated by relative water content were reported for pea (*Pisum sativum*) by Iglesias-García et al. (2015).

A drought tolerance breeding strategy could be first based on laboratory screening of large collections of genetic material for the presence of the identified markers. Then, lines carrying alleles linked to QTLs of targeted traits could be evaluated under field conditions to finally identify drought-tolerant individuals. More focus should be on QTLs related to root-shoot ratio, LRN, SRL, and WS shown to be flanked by SNPs markers. However, the QTL related to WS needs further evaluation under different watering conditions and drought intensity to determine environments of expression of this QTL. This will allow to determine whether it is an adaptive or constitutive QTL.



It should be pointed out that results of QTL analysis using the second round map of JoinMap®4 program (Van Ooijen, 2006) obtained from regression mapping algorithm were closely similar to those obtained using maximum likelihood algorithm, although total lengths of the two maps were different (data not shown).

CONCLUSIONS

In this study, a total of 18 QTLs related to root and shoot traits associated with drought tolerance such as dry root biomass, LRN, root-shoot ratio, and specific root length were identified under progressive drought-stressed treatment. Interestingly, 12 of these QTLs were detected for both seasons, confirming their potential importance in conveying drought tolerance. DNA markers linked to these QTLs could be used for marker-assisted selection, thus making subsequent breeding efforts more reliable and efficient as the respective phenotyping-based methods are slow and labor-intensive, and affected by environment. Although, root characteristics are

difficult to study as many environmental effects (especially soil characteristics) interact with genetic factors, our results provide significant information about QTLs related to root and shoot traits that could be used in marker-assisted breeding after validation.

AUTHOR CONTRIBUTIONS

OI designed the study, analyzed data, interpreted results and wrote the paper. OI, JDR, PVD, EDK, SU contributed to design the study, analyze data, interpret results and wrote the paper. OI, CC, RM, GS contributed to data acquisition. SU, FM contributed to critically revising the paper.

ACKNOWLEDGMENTS

OI and SU thank OCP-Foundation-INRA-ICARDA-IAV Hassan II project on India-Morocco Food Legume Initiative for the support.

REFERENCES

- Abramoff, M. D., Magalhaes, P. J., and Ram, S. J. (2004). Image processing with Image J. *Biophotonics. Inter.* 11, 36–42.
- Arumuganathan, K., and Earle, E. D. (1991). Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* 9, 208–218. doi: 10.1007/BF02672069
- Aswaf, A., and Blair, M. (2012). Quantitative trait loci for rooting pattern traits of common beans grown under drought stress versus non-stress conditions. *Mol. Breed.* 30, 681–695. doi: 10.1007/s11032-011-9654-y
- Ates, D., Sever, T., Aldemir, S., Yagmur, B., Temel, H. Y., Kaya, H. B., et al. (2016). Identification of QTLs controlling genes for Se uptake in lentil seeds. *PLoS ONE* 11:e0149210. doi: 10.1371/journal.pone.0149210
- Brensha, W., Kantartzi, S. K., Meksem, K., Grier, R. L., Barakat, A., Lightfoot, D. A., et al. (2012). Genetic analysis of root and shoot traits in the “Essex” by “Forrest” recombinant inbred line (RIL) population of soybean (*Glycine max* (L.) Merr.). *J. Plant Genome Sci.* 1, 1–9. doi: 10.5147/jpgs.2012.0051
- Cartwright, D. A., Troggio, M., Velasco, R., and Gutin, A. (2007). Genetic mapping in the presence of genotyping errors. *Genetics* 176, 2521–2527. doi: 10.1534/genetics.106.063982
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354
- Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait, J. H. (2011). Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes Genom. Genet.* 1, 171–182. doi: 10.1534/g3.111.000240
- Chen, W., Zhang, Y., Liu, X., Chen, B., Tu, J., and Tingdong, F. (2007). Detection of QTL for six yield-related traits in oilseed rape (*Brassica napus*) using DH and immortalized F(2) populations. *Theor. Appl. Genet.* 115, 849–858. doi: 10.1007/s00122-007-0613-2
- Cichy, K., Blair, M. W., Mendoza, C. H. G., Snapp, S. S., and Kelly, J. D. (2009). QTL analysis of root architecture traits and low phosphorus tolerance in an Andean bean population. *Crop Sci.* 49, 59–68. doi: 10.2135/cropsci2008.03.0142
- Comas, L. H., Becker, S. R., Cruz, V. M. V., Byrne, P. F., and Dierig, D. A. (2013). Root traits contributing to plant productivity under drought. *Front. Plant Sci.* 4:442. doi: 10.3389/fpls.2013.00442
- Day, D. (1991). *Grower Digest, Vol. 12*. London: Growing in Perlite. Grower Publications Ltd.
- De Keyser, E., Yan Shu, Q., Van Bockstaele, E., and De Riek, J. (2010). Multipoint-likelihood maximization mapping on 4 segregating populations to achieve an integrated framework map for QTL analysis in pot azalea (*Rhododendron simsii* hybrids). *BMC Mol. Biol.* 11:1. doi: 10.1186/1471-2199-11-1
- De Riek, J., Calsyn, E., Everaert, I., Van Bockstaele, E., and De Loose, M. (2001). AFLP based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. *Theor. Appl. Genet.* 103, 1254–1265. doi: 10.1007/s001220100710
- Duran, Y., and Perez De La Vega, M. (2004). Assessment of genetic variation and species relationships in a collection of Lens using RAPD and ISSR. *Span J. Agric. Res.* 2, 538–544. doi: 10.5424/sjar/2004024-110
- Eujayl, I., Baum, M., Powell, W., Erskine, W., and Pehu, E. (1998). A genetic linkage map of lentil (*Lens* sp.) based on RAPD and AFLP markers using recombinant inbred lines. *Theor. Appl. Genet.* 97, 83–89. doi: 10.1007/s001220050869
- FAO (2015). *International Year of Pulses 2016*. Available online at: <http://www.fao.org/pulses-2016/en/> (Accessed on: 24 Nov 2015).
- Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Scarascia Mugnozza, G., and Sabatti, M. (2007). Genetic linkage maps of *Populus nigra* L. including AFLPs, SSRs, SNPs, and sex trait. *Tree Genet. Genom.* 4, 25–36. doi: 10.1007/s11295-007-0085-1
- Gaur, P. M., Krishnamurthy, L., and Kashiwagi, J. (2008). Improving drought avoidance root traits in chickpea (*Cicer arietinum*)—current status of research at ICRISAT. *Plant Prod. Sci.* 11, 3–11. doi: 10.1626/pp.s11.3
- Grusak, M. A., and Coyne, C. J. (2009). “Variation for seed minerals and protein concentrations in diverse germplasm of lentil” in *Paper Presented at North America Pulse Improvement Association, 20th Biennial Meeting* (Fort Collins, CO), 11.
- Gupta, M., Verma, B., Kumar, N., Chahota, R. K., Rathour, R., Sharma, S. K., et al. (2012). Construction of intersubspecific molecular genetic map of lentil based on ISSR, RAPD and SSR markers. *J. Genet.* 91, 279–287. doi: 10.1007/s12041-012-0180-4
- Hamwiah, A., Udupa, S. M., Choumane, W., Sarker, A., Dreyer, F., Jung, C., et al. (2005). A genetic linkage map of *Lens* sp. based on microsatellite and AFLP markers and the localization of fusarium vascular wilt resistance. *Theor. Appl. Genet.* 110, 669–677. doi: 10.1007/s00122-004-1892-5
- Idrissi, O., Houasli, C. H., Udupa, S. M., De Keyser, E., Van Damme, P., and De Riek, J. (2015a). Genetic variability for root and shoot traits in a lentil (*Lens culinaris* Medik.) recombinant inbred line population and their association with drought tolerance. *Euphytica* 204, 693–709. doi: 10.1007/s10681-015-1373-8
- Idrissi, O., Udupa, S. M., De Keyser, E., Van Damme, P., and De Riek, J. (2015b). Functional genetic diversity analysis and identification of associated simple sequence repeats and amplified fragment length polymorphism markers to drought tolerance in lentil (*Lens culinaris* ssp. *culinaris* Medicus) Landraces. *Plant Mol. Biol. Rep.* 34, 659–680. doi: 10.1007/s11105-015-0940-4
- Idrissi, O., Udupa, S. M., Houasli, C. H., De Keyser, E., Van Damme, P., and De Riek, J. (2015c). Genetic diversity analysis of Moroccan lentil (*Lens culinaris* Medik.) landraces using simple sequence repeat and amplified fragment length polymorphisms reveals functional adaptation toward agro-environmental origins. *Plant Breed.* 134, 322–332. doi: 10.1111/pbr.12261
- Iglesias-García, R., Prats, E., Fondevilla, S., Satovic, Z., and Rubiales, D. (2015). Quantitative trait loci associated to drought adaptation in pea (*Pisum sativum* L.). *Plant Mol. Biol. Rep.* 33, 1768–1778. doi: 10.1007/s11105-015-0872-z
- Kashiwagi, J., Krishnamurthy, L., Gaur, P. M., Shandra, S., and Upadhyaya, H. D. (2014). Estimation of gene effects of the drought avoidance root characteristics in chickpea (*C. arietinum* L.). *Field Crop Res.* 105, 64–69. doi: 10.1007/s00122-013-2230-6
- Kashiwagi, J., Krishnamurthy, L., Upadhyaya, H. D., Krishna, H., Chandra, S., Vadez, V., et al. (2005). Genetic variability of drought-avoidance root traits in the mini-core germplasm collection of chickpea (*Cicer arietinum* L.). *Euphytica* 146, 213–222. doi: 10.1007/s10681-005-9007-1
- Kaur, S., Cogan, N. O. I., Stephens, A., Noy, D., Butsch, M., Forster, J. W., et al. (2014). EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance. *Theor. Appl. Genet.* 127, 703–713. doi: 10.1007/s00122-013-2252-0
- Kim, S. (2007). Mapping unexplored genome: a genetic linkage map of the woody Sonchus alliance (*Asteraceae*: Sonchineae) in the Macaronesian islands. *J. Hered.* 98, 293–299. doi: 10.1093/jhered/esm052
- Li, G., and Quiros, C. F. (2001). Sequence-related amplified polymorphism (SRAP), a new marker system based on a simple PCR reaction: its application to mapping and gene tagging in *Brassica*. *Theor. Appl. Genet.* 103, 455–461. doi: 10.1007/s001220100570
- Lincoln, S. E., and Lander, E. (1992). Systematic detection of errors in genetic linkage data. *Genomics* 14, 604–610. doi: 10.1016/S0888-7543(05)80158-2
- Lobet, G., Pagès, L., and Draye, X. (2011). A novel image analysis toolbox enabling quantitative analyses of root system architecture. *Plant Physiol.* 157, 29–39. doi: 10.1104/pp.111.179895
- Malhotra, R. S., Sarker, A., and Saxena, M. C. (2004). “Drought tolerance in chickpea and lentil—present status and future strategies,” in *Challenges and Strategies for Dryland Agriculture*, eds S. C. Rao and J. Ryan (Wisconsin, EI: CSSA Special Publication no. 32, Crop Science Society of America and American Society of Agronomy), 257–273.
- Manavalan, L. P., Prince, S. J., Musket, T. A., Chaky, J., Deshmukh, R., Vuong, T. D., et al. (2015). Identification of novel QTL governing root architectural traits in an interspecific soybean population. *PLoS ONE* 10:e0120490. doi: 10.1371/journal.pone.0120490
- Muys, C., Noelle, C., Hienpont, T., Dauchot, N., Maudoux, O., Draye, X., et al. (2014). Integration of AFLPs, SSRs and SNPs markers into a new genetic map of industrial chicory (*Cichorium Intybus* L. var. *sativum*). *Plant Breed.* 133, 130–137. doi: 10.1111/pbr.12113
- Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. (2012). Development of high density maps for barley and wheat using a novel two-enzyme genotyping by sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253
- Rabah Nasser, R. (2009). *Physiological Aspects of the Response to Elevated CO₂ in Lentils (Lens culinaris Medic.)*. Dissertation thesis: School of Biological Sciences of the University of Plymouth, Devon.

- Roberts, D. W. H., and Wolfe, A. D. (2014). Sequence-Related Amplified Polymorphism (SRAP) Markers: a potential resource for studies in plant molecular biology. *Appl. Plant Sci.* 2:1400017. doi: 10.3732/apps.1400017
- Rubeena, Ford, R., and Taylor, P. W. (2003). Construction of an intraspecific linkage map of lentil (*Lens culinaris* ssp. *culinaris*). *Theor. Appl. Genet.* 107, 910–916. doi: 10.1007/s00122-003-1326-9
- Ruta, N. (2008). *Quantitative Trait Loci Controlling Root and Shoot Traits of Maize under Drought Stress*. Dissertation thesis, Swiss Federal Institute of Technology Zurich, Zurich.
- Saha, G. C., Sarker, A., Chen, W., Vandemark, G. J., and Muehlbauer, F. J. (2010). Inheritance and linkage map positions of genes conferring resistance to Stemphylium Blight in lentil. *Crop Sci.* 50, 1831–1839. doi: 10.2135/cropsci2009.12.0709
- Saha, G. C., Sarker, A., Chen, W., Vandemark, G. J., and Muehlbauer, F. J. (2013). Inheritance and linkage map positions of genes conferring agromorphological traits in *Lens culinaris* Medik. *Int. J. Agron.* 2013. doi: 10.1155/2013/618926
- Sarker, A., Aydogan, A., Chandra, S., Kharrat, M., and Sabaghpour, S. (2009). “Genetic enhancement for yield and yield stability,” in *The Lentil: Botany, Production and Uses*, eds W. Erskine, F. J. Muehlbauer, A. Sarker, and B. Sharma (Oxfordshire: CAB International), 102–120.
- Sarker, A., Erskine, W., and Singh, M. (2005). Variation in shoot and root characteristics and their association with drought tolerance in lentil landraces. *Genet. Resour. Crop. Evol.* 52, 89–97. doi: 10.1007/s10722-005-0289-x
- Sayed, M. A. A. (2011). *QTL Analysis for Drought Tolerance Related to Root and Shoot Traits in Barley (Hordeum vulgare L.)*. Dissertation thesis, Institut für Nutzpflanzenwissenschaften und Ressourcenschutz, Hohen Landwirtschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn, Bonn.
- Sharpe, A. G., Ramsay, L., Sanderson, L. A., Fedoruk, M. J., Clarke, W. E., Li, R., et al. (2013). Ancient orphan crop joins modern era: gene-based SNP discovery and mapping in lentil. *BMC Genomics* 14:192. doi: 10.1186/1471-2164-14-192
- Singh, D. H., Dikshit, H. K., and Singh, R. A. (2013). A new phenotyping technique for screening for drought tolerance in lentil (*Lens culinaris* Medik.). *Plant Breed.* 132, 185–190. doi: 10.1111/pbr.12033
- Stoddard, F. L., Balko, C., Erskine, W., Khan, H. R., Link, W., and Sarker, A. (2006). Screening techniques and sources of resistance to abiotic stresses in cool-season food legumes. *Euphytica* 147, 167–186. doi: 10.1007/s10681-006-4723-8
- Thavarajah, D., Thavarajah, P., Sarker, A., Materne, M., Vandemark, G., Shrestha, R., et al. (2011). A global survey of effects of genotype and environment on selenium concentration in lentils (*Lens culinaris* L.): implications for nutritional fortification strategies. *Food Chem.* 125, 72–76. doi: 10.1016/j.foodchem.2010.08.038
- Ting, N. C., Jansen, J., Mayes, S., Massawe, F., Sambanthamurthi, R., Ooi, L. C. L., et al. (2014). High density SNP and SSR-based genetic maps of two independent oil palm hybrids. *BMC Genomics* 15:309. doi: 10.1186/1471-2164-15-309
- Vadez, V., Rao, S., Kholova, J., Krishnamurthy, L., Kashiwaji, J., Ratnakumar, P., et al. (2008). Root research for drought tolerance in legumes: quo-vadis? *J. Food Legumes* 21, 77–85.
- Van Ooijen, J. W. (2004). *MapQTL® 5, Software for the Mapping of Quantitative Trait Loci in Experimental Populations*. Wageningen: Kyasma B.V.
- Van Ooijen, J. W. (2006). *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyasma B.V.
- Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwaji, J., Krishnamurthy, L., et al. (2014). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6
- Verslues, P. E., Agarwal, M., Agarwal, S. K., Zhu, J., and Zhu, J. K. (2006). Methods and concepts in quantifying resistance to drought, salt and freezing, abiotic stresses that affect plant water status. *Plant J.* 45, 523–539. doi: 10.1111/j.1365-3113X.2005.02593.x
- Voorrips, R. E. (2002). MapChart®: Software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Vos, P., Hogers, R., Bleeker, M., Reijans, M., Van de Lee, T., Hornes, M., et al. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* 23, 4407–4414. doi: 10.1093/nar/23.21.4407
- Wong, M. M., Gujaria-Verma, N., Ramsay, L., Yuan, H. Y., Caron, C., Diapari, M., et al. (2015). Classification and characterization of species within the genus *lens* using genotyping-by-sequencing (GBS). *PLoS ONE* 27:e0122025. doi: 10.1371/journal.pone.0122025
- Yuan, X. J., Li, X. Z., Pan, J. S., Wang, G., Jiang, S., Li, X. H., et al. (2008). Genetic linkage map construction and location of QTLs for fruit-related traits in cucumber. *Plant Breed.* 127, 180–188. doi: 10.1111/j.1439-0523.2007.01426.x
- Zhang, Z., Hu, M., Zhang, J., Liu, D., Zheng, J., Zhang, K., et al. (2009). Construction of a comprehensive PCR-based marker linkage map and QTL mapping for fiber quality traits in upland cotton (*Gossypium hirsutum* L.). *Mol. Breed.* 24, 49–61. doi: 10.1007/s11032-009-9271-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Idrissi, Udupa, De Keyser, McGee, Coyne, Saha, Muehlbauer, Van Damme and De Riek. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Polyamines Confer Salt Tolerance in Mung Bean (*Vigna radiata* L.) by Reducing Sodium Uptake, Improving Nutrient Homeostasis, Antioxidant Defense, and Methylglyoxal Detoxification Systems

Kamrun Nahar^{1,2}, Mirza Hasanuzzaman^{3*}, Anisur Rahman^{1,3}, Md. Mahabub Alam¹, Jubayer-Al Mahmud^{1,4}, Toshisada Suzuki⁵ and Masayuki Fujita¹

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica – Universidade Nova
de Lisboa, Portugal

Reviewed by:

Li-Song Chen,
Fujian Agriculture and Forestry
University, China
Miguel López Gómez,
University of Granada, Spain

*Correspondence:

Mirza Hasanuzzaman
mhzsauag@yahoo.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 09 May 2016

Accepted: 12 July 2016

Published: 28 July 2016

Citation:

Nahar K, Hasanuzzaman M,
Rahman A, Alam MM, Mahmud J-A,
Suzuki T and Fujita M (2016)
Polyamines Confer Salt Tolerance
in Mung Bean (*Vigna radiata* L.) by
Reducing Sodium Uptake, Improving
Nutrient Homeostasis, Antioxidant
Defense, and Methylglyoxal
Detoxification Systems.
Front. Plant Sci. 7:1104.
doi: 10.3389/fpls.2016.01104

¹ Laboratory of Plant Stress Responses, Faculty of Agriculture, Kagawa University, Kagawa, Japan, ² Department of Agricultural Botany, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, ³ Department of Agronomy, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, ⁴ Department of Agroforestry and Environmental Science, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, ⁵ Biomass Chemistry Laboratory, Bioresource Science for Manufacturing, Department of Applied Bioresource Science, Faculty of Agriculture, Kagawa University, Kagawa, Japan

The physiological roles of PAs (putrescine, spermidine, and spermine) were investigated for their ability to confer salt tolerance (200 mM NaCl, 48 h) in mung bean seedlings (*Vigna radiata* L. cv. BARI Mung-2). Salt stress resulted in Na toxicity, decreased K, Ca, Mg, and Zn contents in roots and shoots, and disrupted antioxidant defense system which caused oxidative damage as indicated by increased lipid peroxidation, H₂O₂ content, O₂^{•−} generation rate, and lipoxygenase activity. Salinity-induced methylglyoxal (MG) toxicity was also clearly evident. Salinity decreased leaf chlorophyll (chl) and relative water content (RWC). Supplementation of salt affected seedlings with exogenous PAs enhanced the contents of glutathione and ascorbate, increased activities of antioxidant enzymes (dehydroascorbate reductase, glutathione reductase, catalase, and glutathione peroxidase) and glyoxalase enzyme (glyoxalase II), which reduced salt-induced oxidative stress and MG toxicity, respectively. Exogenous PAs reduced cellular Na content and maintained nutrient homeostasis and modulated endogenous PAs levels in salt affected mung bean seedlings. The overall salt tolerance was reflected through improved tissue water and chl content, and better seedling growth.

Keywords: abiotic stress, salinity, polyamine, methylglyoxal, oxidative damage, ROS signaling

Abbreviations: AO, ascorbate oxidase; APX, ascorbate peroxidase; AsA, ascorbic acid (ascorbate); BSA, bovine serum albumin; CAT, catalase; CDNB, 1-chloro-2, 4-dinitrobenzene; chl, chlorophyll; DHA, dehydroascorbate; DHAR, dehydroascorbate reductase; DTNB, 5,5'-dithio-bis (2-nitrobenzoic acid); EDTA, ethylenediaminetetraacetic acid; Gly I, glyoxalase I; Gly II, glyoxalase II; GR, glutathione reductase; GSH, reduced glutathione; GSSG, oxidized glutathione; GPX, glutathione peroxidase; GST, glutathione S-transferase; LOX, lipoxygenase; MDA, malondialdehyde; MDHA, monodehydroascorbate; MDHAR, monodehydroascorbate reductase; MG, methylglyoxal; NADPH, nicotinamide adenine dinucleotide phosphate; NTB, 2-nitro-5-thiobenzoic acid; Pro, Proline; ROS, reactive oxygen species; RWC, relative water content; SLG, S-D-lactoylglutathione; SOD, superoxide dismutase; TBA, thiobarbituric acid; TCA, trichloroacetic acid.

INTRODUCTION

Plants show susceptibility to various stresses under both wild and cultivated conditions. In many areas of the world, high concentrations of salt in the soil are commonly encountered phenomenon. Salinity is a complex stress that involves both ionic and osmotic components. Salt stress adversely affects plant development and productivity by generating ion toxicity, inducing nutritional deficiencies, creating osmotic stress and water deficits, and decreasing photosynthesis (Shabala and Munns, 2012).

Reactive oxygen species are partly reduced forms of oxygen and by-products of aerobic metabolism and stronger oxidants than molecular oxygen. ROS are overproduced under stress condition and can potentially damage the intracellular machinery. ROS may include free radicals such as superoxide anion ($O_2^{\bullet-}$) and hydroxyl radical ($\bullet OH$), as well as non-radical molecules like hydrogen peroxide (H_2O_2) and singlet oxygen (1O_2). Plants can scavenge a certain amount of ROS generated as by-products of aerobic metabolism, but under stress conditions, including salt stress, the resulting photoinhibition often creates an imbalance between ROS generation and scavenging due to generation of excessive amount of ROS, thereby resulting in oxidative stress, causing peroxidation of lipids, oxidation of proteins, inhibition of enzyme activity, injury to nucleic acids, activation of programmed cell death (PCD) pathways, and eventually leading to death of the cells (Gill and Tuteja, 2010; Hasanuzzaman et al., 2013). The antioxidant defense machinery is a ROS scavenging system, which protects plants from oxidative damage. Efficient enzymatic (superoxide dismutase, SOD; catalase, CAT; ascorbate peroxidase, APX; glutathione reductase, GR; monodehydroascorbate reductase, MDHAR; dehydroascorbate reductase, DHAR; glutathione peroxidase, GPX; guaiacol peroxidase, GOPX; and glutathione S-transferase, GST) and non-enzymatic (ascorbate, AsA; glutathione, GSH; phenolic compounds, alkaloids, non-protein amino acids, and α -tocopherols) components form an antioxidant defense system to protect overproduction of ROS and to prevent plant cells from experiencing oxidative damage (Gill and Tuteja, 2010).

Methylglyoxal (MG) is a transition-state intermediate product of the triose-phosphates of the glycolysis pathway in eukaryotic cells. High accumulation of MG is toxic and inhibits cell proliferation, causes degradation of proteins by modifying arginine, lysine, and cysteine residues, forms adducts with guanyl nucleotide in DNA, and inactivates antioxidant defense system, and causes oxidative stress (Yadav et al., 2005). The glyoxalase system comprises of two enzymes, gly I (glyoxalase I) and gly II (glyoxalase II), which catalyze the detoxification reaction of MG to D-lactate (Yadav et al., 2005). Although the glyoxalase enzymes have been extensively studied in microbial and animal systems (Thornalley, 1990), the biological significance of this pathway in plants and under stress condition is just beginning to be explored (Yadav et al., 2005).

The polyamines (PAs) are low-molecular-weight organic cations found in a wide range of organisms, where they perform diverse biological functions (Kusano et al., 2008). Putrescine (Put), spermidine (Spd), and spermine (Spm) are the most

common PAs. The levels of PAs frequently increase during stress and they play roles in enhancing plant stress tolerance. Acting as molecular chaperones, PAs bind to negatively charged surfaces and protect membranes and biomolecules (Kusano et al., 2008). PAs may also act as ROS and free-radical scavengers and activate the antioxidant enzyme machinery which help to reduce oxidative stress and subsequent membrane injury and electrolyte leakage (Gill and Tuteja, 2010; Kubiś et al., 2014). PAs interact with some signaling molecules (Kusano et al., 2008). PAs have been confirmed to be unique polycationic metabolites involved in the direct blockage of a variety of cation and K^+ -selective channels, vacuolar-type channels, and ammonium channels (Dobrovinskaya et al., 1999). PAs restrict plasma membrane Na^+ influx and $NaCl$ -induced K^+ efflux and shoot-to-root K^+ recirculation (Shabala et al., 2007; Zhao et al., 2007). Despite their range of protective effects under stress condition, acute application of PAs can cause endogenous PAs catabolism in the apoplast, which is responsible for ROS-induced oxidative damage (Di Tomaso et al., 1989). The diversity of the physiological functions of PAs has made understanding the detailed mechanisms of the protective function of PAs in response to stress particularly challenging.

Salt stress tolerance is a polygenic trait. Some of the major traits affected by salinity include the maintenance of membrane transport activity and ionic homeostasis, compartmentalization of ions at the cellular and whole-plant level (which may include Na^+ exclusion from root uptake, intra-cellular Na^+ sequestration, cytosolic K^+ retention), osmotic adjustment through synthesis of compatible solutes, oxidative stress tolerance, biomembrane protection, and MG toxicity tolerance (Yadav et al., 2005; Shabala and Munns, 2012; Rahman et al., 2016). Changes in ranges of biochemical pathways and the induction of hormones and signaling molecules activate stress responsive genes that impart salt stress tolerance (Adem et al., 2014). Polyamines have diverse mechanism to enhance salt tolerance. Under salt stress PAs proved to be efficient in modulating membrane integrity and ion transport processes, vacuolar ion channels, shoot, and root growth (Di Tomaso et al., 1989; Dobrovinskaya et al., 1999), photosynthesis (Kotzabasis et al., 1993; Duan et al., 2008; Shu et al., 2012), antioxidative responses (Sheokand et al., 2008; Iqbal et al., 2014); Na^+ and nutrient influx/efflux (Shabala et al., 2007; Pottosin et al., 2012). However, the actual mechanism by which PAs modulate these physiological processes is not yet known. A few components of antioxidant system were studied in response to PAs modulation. Polyamine-induced MG detoxification process is unknown in plants. In the present study, we have investigated the roles of PAs in maintaining ionic and nutrient homeostasis, inducing osmotic adjustment, and protecting mung bean plant cells from oxidative damage and MG toxicity under salt stress. The results presented here provide new insight into the role of PAs in enhancing salt tolerance and mitigating salt stress-induced damage in mung bean seedlings by the coordinated action of PAs on ion homeostasis, antioxidant defense, and the glyoxalase system.

MATERIALS AND METHODS

Plant Materials, Growth Condition, and Stress Treatments

Mung bean (*Vigna radiata* L. cv. BARI Mung-2) seedlings were grown in petri dishes under the conditions: light, $350 \mu\text{mol photon m}^{-2} \text{ s}^{-1}$; temperature, $25 \pm 2^\circ\text{C}$; relative humidity, 65–70%; 10,000-fold diluted Hyponex solution (Hyponex, Japan) was applied as nutrient. Six-day-old seedlings (three sets) were exposed to salt stress (NaCl, 200 mM). Three sets of 5-day-old seedlings were grown with Put (0.2 mM), Spd (0.2 mM), and Spm (0.2 mM) solution as pre-treatment for 24 h. These pre-treated seedlings were then exposed to the same level of salt stress on day six. Control seedlings were grown with Hyponex solution. Another three sets of seedlings were grown with Put, Spd and Spm without any stress. Data were taken after 48 h. There were three replicates per treatment.

Measurement of Na and Mineral Nutrients Contents in Roots and Shoots

Root and shoot samples were oven-dried at 80°C for 48 h. Dried samples were ground and subjected to acid digestion in $\text{HNO}_3\text{:HClO}_4$ (5:1 v/v) mixture at 80°C . The Na, K, Ca, Mg, and Zn contents were measured using flame atomic absorption spectrophotometer.

Histochemical Detection of Hydrogen Peroxide and Superoxide

The $\text{HO}_2^{\bullet-}$ and $\text{O}_2^{\bullet-}$ were localized histochemically (Chen et al., 2010) by staining leaves with 1% 3,3-diaminobenzidine (DAB) and 0.1% nitroblue tetrazolium chloride (NBT) solution, respectively.

Lipid Peroxidation, H_2O_2 Content and $\text{O}_2^{\bullet-}$ Generation Rate

The level of lipid peroxidation was measured by estimating (MDA, a product of lipid peroxidation) using TBA according to Heath and Packer (1968) with modifications (Hasanuzzaman et al., 2011). Hydrogen peroxide (H_2O_2) was assayed according to Yu et al. (2003) by extracting leaves in potassium-phosphate buffer (pH 6.5) (centrifuging at $11,500 \times g$), then adding it to a mixture of TiCl_4 in 20% H_2SO_4 (v/v) and the resulting solution was measured spectrophotometrically at 410 nm. To measure rate of $\text{O}_2^{\bullet-}$ generation leaves were homogenized in K-P buffer solution, centrifuged at $5000 \times g$. Supernatant was mixed with extraction buffer and hydroxylamine hydrochloride for incubation; then mixed with sulfanilamide and naphthylamine, incubated at 25°C . The absorbance was measured at 530 nm (Nahar et al., 2016).

Extraction and Measurement of Ascorbate and Glutathione

Leaves (0.5 g) were homogenized in 5% meta-phosphoric acid containing 1 mM EDTA (centrifuged at $11,500 \times g$) for 15 min at 4°C and the supernatant was collected for analysis of ascorbate

and glutathione. Ascorbate content was determined following the method of Hasanuzzaman et al. (2011). To determine total ascorbate, the oxidized fraction was reduced by adding 0.1 M dithiothreitol for 1 h at room temperature and then read at 265 nm using 1.0 unit AO. Oxidized ascorbate (DHA) content was determined by subtracting reduced AsA from total AsA. The glutathione pool was assayed according to previously described methods (Yu et al., 2003) with modifications as described by Hasanuzzaman et al. (2011). Standard curves with known concentrations of GSH and GSSG were used. The content of GSH was calculated by subtracting GSSG from total GSH.

Protein Determination and Enzyme Extraction and Assays

The protein concentration was determined according to Bradford (1976). Leaves were homogenized with 50 mM K-P (potassium phosphate) buffer (pH 7.0) containing 100 mM KCl, 1 mM AsA, 5 mM β -mercaptoethanol, and 10% (w/v) glycerol, centrifuged at $11,500 \times g$ and supernatants were used for enzyme activity assay. SOD (EC 1.15.1.1) activity (El-Shabrawi et al., 2010), CAT (EC: 1.11.1.6) activity (Hasanuzzaman et al., 2011), APX (EC: 1.11.1.11) (Nakano and Asada, 1981), MDHAR (EC: 1.6.5.4) (Hossain et al., 1984), DHAR (EC: 1.8.5.1) (Nakano and Asada, 1981), GR (EC: 1.6.4.2) (Hasanuzzaman et al., 2011), and GST (EC: 2.5.1.18) (Hossain et al., 2006) activities were measured following standard methodologies. GPX (EC: 1.11.1.9), Gly I (EC: 4.4.1.5), Gly II (EC: 3.1.2.6) activity was measured following Hasanuzzaman et al. (2011). LOX (EC 1.13.11.12) activity was estimated as described in Nahar et al. (2016).

Methylglyoxal Level

Leaves were homogenized in 5% perchloric acid and centrifuged at $11,000 \times g$, supernatant was decolorized, neutralized, MG content was estimation by adding sodium dihydrogen phosphate and N-acetyl-L-cysteine to a final volume of 1 mL. Formation of the product N- α -acetyl-S-(1-hydroxy-2-oxo-prop-1-yl)cysteine was recorded after 10 min at a wavelength of 288 nm (Wild et al., 2012).

Measurement of Free Polyamine Content

Endogenous free PAs were estimated according to Kotzabasis et al. (1993). In brief, leaf tissue (0.1 g) was homogenized in 1 mL of 5% (v/v) cold perchloric acid (PCA). The homogenates were kept at 2°C for 2 h and centrifuged at $15,000 \times g$ for 20 min. The supernatant was collected and stored at 2°C . PA standards, free PAs were benzoylated. Briefly, 1 ml of 2N NaOH and 10 μl of benzoylchloride were added to 200 μl of the PA aliquots and vortexed for 30 s. After 20 min incubation at 25°C , 2 ml of saturated NaCl was added to stop the reaction. The benzoyl – PAs were gently mixed with 2 ml diethyl ether. After properly vortex, upper portion was taken and the ether phase was collected and evaporated to dryness in a water bath (60°C). The benzoyl-PAs were redissolved in 200 μl of 64% (v/v) methanol and analyzed by HPLC. A 20 μl of the benzoylated extract was injected into the C18 reverse phase HPLC column (4.6 mm \times 100 mm, 5 μm particle size,

Inertsil.ODS-3). The mobile phase was methanol and water in the ratio of 64:36, respectively, at a isocratic flow rate of 1.0 ml min^{-1} and peaks were detected with a UV detector at 254 nm. Three polyamine standards of Put, Spd and Spm were prepared at different concentrations for the production of standard curves. The final PAs content was expressed as $\mu\text{mol g}^{-1} \text{ DW}$.

Leaf Relative Water Content

Leaf relative water content (RWC) of leaf was measured according to Barrs and Weatherley (1962). Fresh weight (FW), turgid weight (TW), and dry weight (DW) of leaves were measured, and RWC was calculated using the following formula: $\text{RWC (\%)} = [(FW-DW)/(TW-DW)] \times 100$.

Proline Content

Proline (Pro) was calculated according to Bates et al. (1973). Leaves were homogenized in 3% sulphosalicylic acid and centrifuged at $11,500 \times g$. Supernatant was mixed with acid ninhydrin with glacial acetic acid and phosphoric acid. After incubating the mixture at 100°C for 1 h and cooling, toluene was added; chromophore containing toluene was read spectrophotometrically at 520 nm.

Chlorophyll Content

Leaf supernatant was extracted with 80% v/v acetone (centrifuging at $5,000 \times g$), absorbances were taken with a UV-visible spectrophotometer at 663 and 645 nm for chl *a* and chl *b*, respectively, and chl content was calculated according to Arnon (1949).

Determination of Growth Parameters

Plant height and root length were measured from each set of seedlings. Ten randomly selected fresh seedlings from each treatment were dried at 80°C for 48 h, then weighed and considered as DW.

Statistical Analysis

All data obtained were subjected to analysis of variance (ANOVA) and the mean differences were compared by Tukey's HSD (honest significant difference) test using XLSTAT v. 2015.1.01 software (Addinsoft, 2015). Differences at $P \leq 0.05$ were considered significant.

RESULTS

Exogenous PAs Reduces Na Uptake

The protective role of PAs against salt stress was examined by determining the Na content in roots and shoots of the mung bean seedlings. As shown in **Figures 1A,B**, salt treatment resulted in a marked increase in Na contents in the roots and shoots when compared to control seedlings. However, the application of exogenous PAs significantly decreased Na level in root and shoot (**Figures 1A,B**).

Mineral Nutrient Contents Induced by Exogenous PAs under Salt Stress

The levels of K in the roots and shoots, decreased significantly under NaCl stress. However, the application of exogenous PAs significantly increased K level in the roots and shoots exposed to salt stress, in contrast to salt treatment without PAs (**Figures 1C,D**). Root and shoot Ca content decreased by 63 and 71%, respectively; Mg content by 24 and 39%, respectively; and Zn content by 48 and 21%, respectively in salt affected mung bean seedlings when compared to control seedlings. Exogenous supplementation of PAs together with the salt treatment increased Ca content in root and Mg content in roots and shoots, compared to the salt treatment alone. Polyamine application with salt stress slightly increased the Zn level in roots and shoots, but the difference was not statistically significant (**Figures 1E-J**).

Histochemical Detection of ROS

The accumulations of the ROS, H_2O_2 , and $\text{O}_2^{\bullet-}$, were detected by histochemical staining with DAB or NBT, respectively. Both H_2O_2 and $\text{O}_2^{\bullet-}$ staining is clearly observed in leaves as brown patches and dark blue spots, respectively. Leaves of salt-treated mung bean seedlings had more prominent and frequent spots. Exogenous PAs reduced the numbers of spots due to H_2O_2 and $\text{O}_2^{\bullet-}$ in the leaves of salt-treated plants (**Figures 2A,B**).

Reactive Oxygen Species and Oxidative Stress Marker

Salt stress results in a ROS burst in mung bean seedlings. Thus, compared to control, high increases in H_2O_2 content and $\text{O}_2^{\bullet-}$ generation rate were evident in salt-affected seedlings. The increase in MDA/lipid peroxidation, indicating damage to membranes, paralleled the accumulation of ROS (H_2O_2 and $\text{O}_2^{\bullet-}$) and the increased activity of lipid degrading enzymes/LOX (**Figures 2C-F**).

Exogenous PAs reduced ROS accumulation in the leaves. Put, Spd, and Spm supplementation during salt stress reduced H_2O_2 content by 23, 21, and 21%, respectively, and reduced $\text{O}_2^{\bullet-}$ generation rate by 27, 27, and 20%, respectively (compared to salt affected seedlings without PA application). LOX activity was also decreased by PAs. Lipid peroxidation or MDA content was reduced by 26, 35, and 39% after Put, Spd, and Spm supplementation of salt-stressed plants (compared to salt treatment alone) (**Figures 2C-F**).

Non-enzymatic Antioxidants

Ascorbate content reduced under salt stress, whereas DHA content increased. Salt stress resulted in a decrease in AsA/DHA ratio (**Figures 3A-C**). Glutathione and GSSG contents increased, but the ratio of GSH/GSSG decreased in response to salt stress, when compared to the control seedlings (**Figures 3D-F**). Exogenous application of Put, Spd, and Spm decreased DHA content and increased content of AsA and the ratio of AsA/DHA (**Figures 3A-C**). Application of PAs also increased GSH content and decreased GSSG content, thereby increasing the ratio of GSH/GSSG (**Figures 3D-F**).

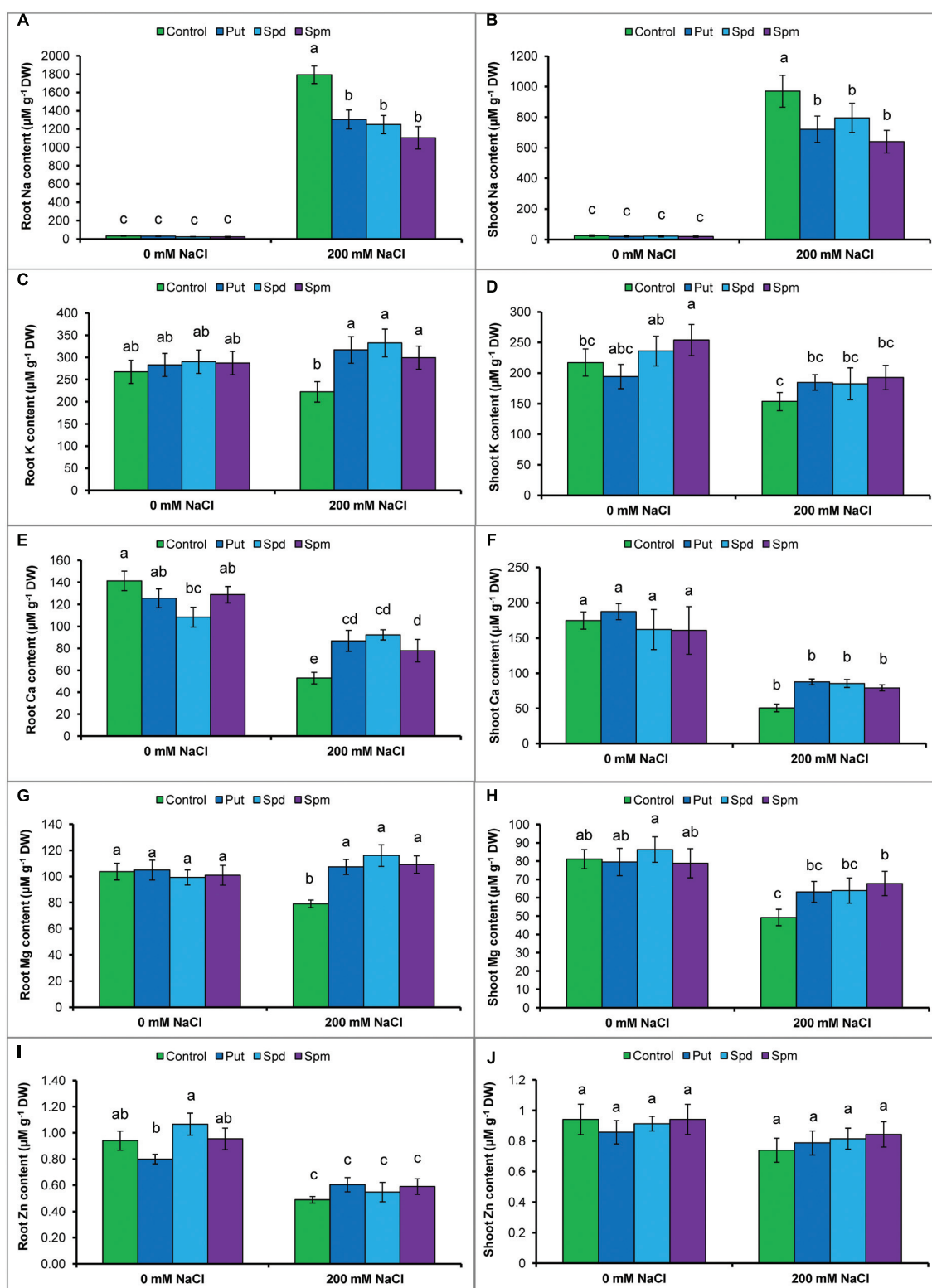


FIGURE 1 | Effect of exogenous polyamines on Na uptake and nutrient homeostasis in mung bean seedlings under salt stress (NaCl, 200 mM). Root Na (A), shoot Na (B), root K (C), and shoot K (D) root Ca (E), shoot Ca (F), root Mg (G), shoot Mg (H), root Zn (I), and shoot Zn (J) contents. Here, Put, Spd, and Spm indicate putrescine (0.2 mM), spermidine (0.2 mM), and spermine (0.2 mM), respectively. Mean (\pm SD) was calculated from three replicates for each treatment. Bars with different letters are significantly different at $P \leq 0.05$ applying Tukey's HSD test.

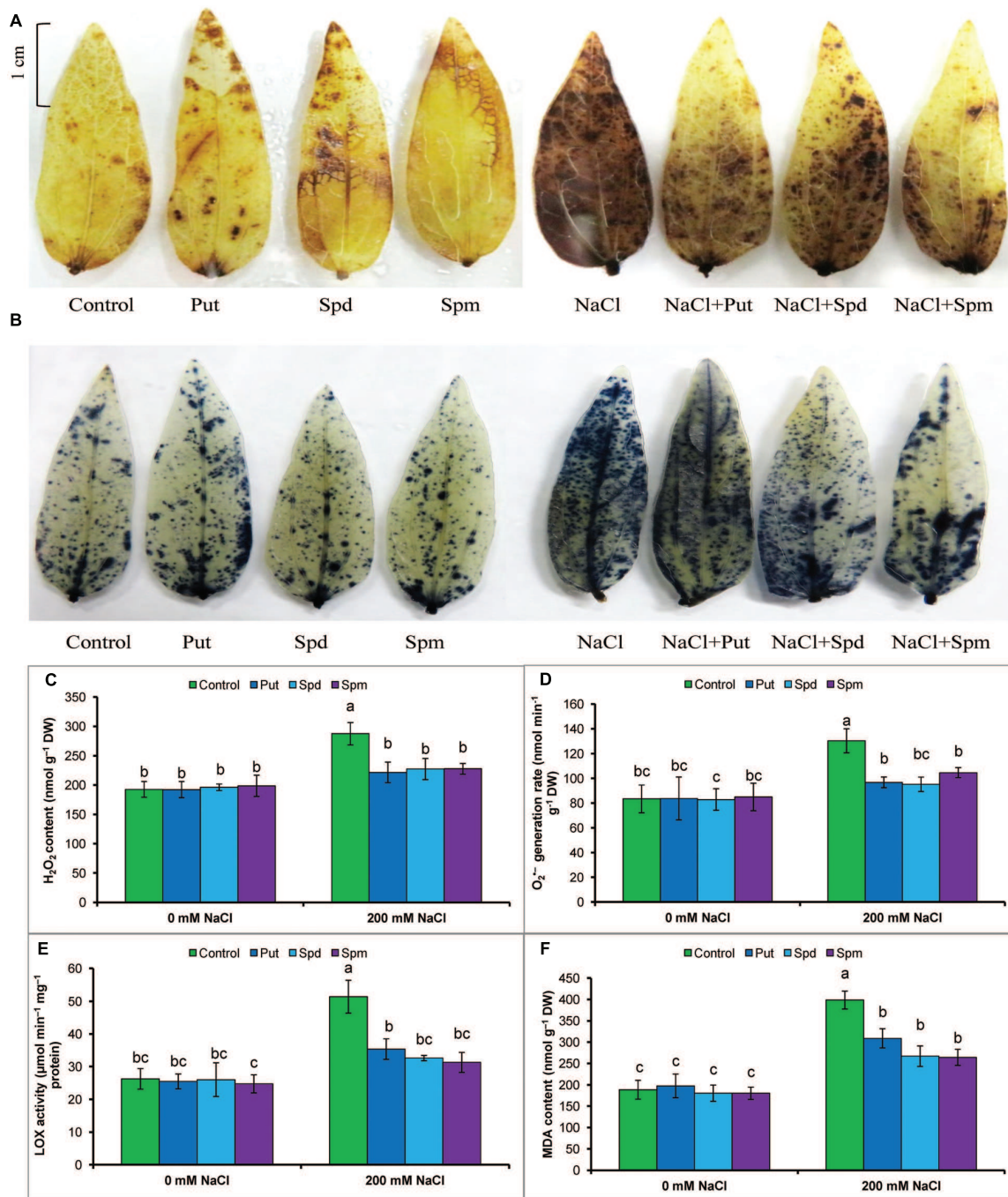


FIGURE 2 | Histochemical detection of ROS in leaves, content of ROS, LOX activity, and membrane lipid peroxidation in salt (NaCl, 200 mM) affected mung bean seedlings. DAB staining (A) of H₂O₂ and NBT staining (B) of O₂⁻ in leaves, H₂O₂ content (C), O₂⁻ generation rate (D), LOX activity (E), and MDA content (F). Here, Put, Spd, and Spm indicate putrescine (0.2 mM), spermidine (0.2 mM), and spermine (0.2 mM), respectively. Mean (±SD) was calculated from three replicates for each treatment. Bars with different letters are significantly different at $P \leq 0.05$ applying Tukey's HSD test.

Antioxidants Enzymes

The activity of SOD increased by 49% in plants under salt stress compared to control seedlings. Exogenous PAs application to the salt affected seedlings did not increase SOD activity further but maintained the same activity as seen in salt affected

seedlings (Table 1). CAT activity decreased by 50% due to salt stress compared to control seedlings. Exogenous Put, Spd, and Spm restored and increased CAT activity in salt stressed mung bean seedlings (compared to seedlings treated with salt only) (Table 1).

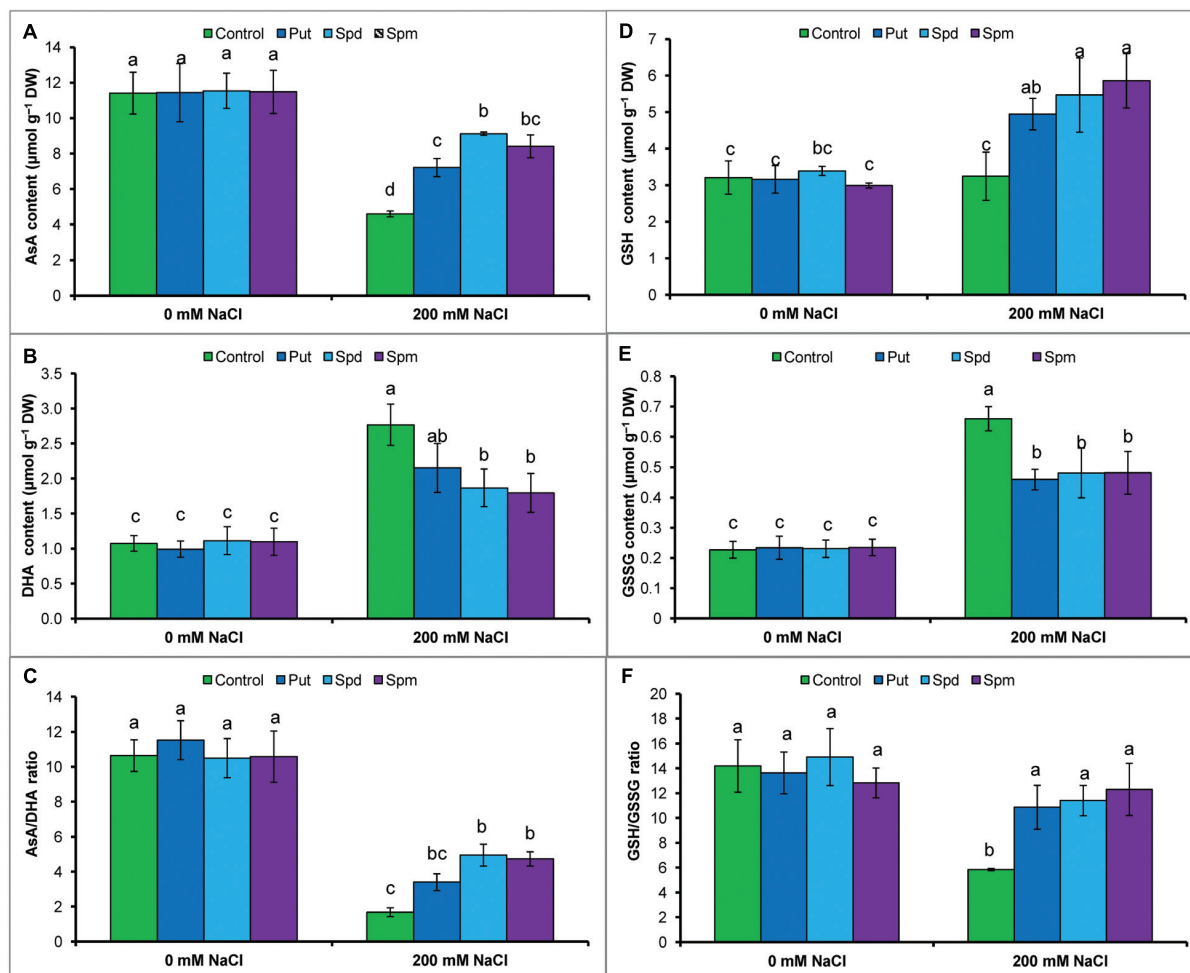


FIGURE 3 | Effect of exogenous polyamines on non-enzymatic antioxidants in mung bean seedlings under salt (NaCl, 200 mM) stress. Ascorbate (AsA) content (A), DHA content (B), AsA/DHA ratio (C), GSH content (D), GSSG content (E), and GSH/GSSG ratio (F). Here, Put, Spd, and Spm indicate putrescine (0.2 mM), spermidine (0.2 mM), and spermine (0.2 mM), respectively. Mean (\pm SD) was calculated from three replicates for each treatment. Bars with different letters are significantly different at $P \leq 0.05$ applying Tukey's HSD test.

The activity of APX increased under salt stress (compared to control). In contrast to the control, salt affected mung bean seedlings showed reductions in MDHAR and DHAR activities. GR activity remained unchanged in salt stressed seedlings, compared to control plants. Activities of APX and MDHAR did not show further increases in response to the application of exogenous PAs with salt stress (when compared to plants treated with salt only). Exogenous PAs application with salt stress enhanced activities of DHAR and GR, when compared to plants exposed to the salt treatment alone (Table 1).

The activity of GST increased by 88% under salt stress compared to control plants. When compared to salt stress alone, the addition of PAs to salt stressed plants resulted in the same GST activity (Table 1). Salt stress did not significantly affect the GPX activity, when compared to control plants. In contrast, salt stressed seedlings supplemented with PAs showed higher GPX activity, when compared to plants treated with salt alone (Table 1).

Methylglyoxal Toxicity and Glyoxalase System

Methylglyoxal content increased by 109% under salt stress compared to control. Activity of Gly I remained unchanged but Gly II reduced by 33% in salt treated seedlings. However, salt affected seedlings supplemented with Put, Spd, and Spm showed enhanced Gly II activity and reduced MG contents (compared to salt stressed seedlings without any exogenous protectant) (Figures 4A–C).

Endogenous Free Polyamines Contents

Mung bean seedlings exposed to salt stress accumulated high levels of endogenous free Put, Spd, and Spm (compared to control). The highest increase was observed for Put content, which reduced the ratio of (Spd+Spm)/Put. Exogenous application of PAs to salt stressed seedlings did not change the Put content, but increased Spd and Spm contents, which restored and enhanced the ratio of (Spd+Spm)/Put (Figures 5A–D).

MATERIALS AND METHODS

Plant Materials, Growth Condition, and Stress Treatments

Mung bean (*Vigna radiata* L. cv. BARI Mung-2) seedlings were grown in petri dishes under the conditions: light, $350 \mu\text{mol photon m}^{-2} \text{ s}^{-1}$; temperature, $25 \pm 2^\circ\text{C}$; relative humidity, 65–70%; 10,000-fold diluted Hyponex solution (Hyponex, Japan) was applied as nutrient. Six-day-old seedlings (three sets) were exposed to salt stress (NaCl, 200 mM). Three sets of 5-day-old seedlings were grown with Put (0.2 mM), Spd (0.2 mM), and Spm (0.2 mM) solution as pre-treatment for 24 h. These pre-treated seedlings were then exposed to the same level of salt stress on day six. Control seedlings were grown with Hyponex solution. Another three sets of seedlings were grown with Put, Spd and Spm without any stress. Data were taken after 48 h. There were three replicates per treatment.

Measurement of Na and Mineral Nutrients Contents in Roots and Shoots

Root and shoot samples were oven-dried at 80°C for 48 h. Dried samples were ground and subjected to acid digestion in $\text{HNO}_3\text{:HClO}_4$ (5:1 v/v) mixture at 80°C . The Na, K, Ca, Mg, and Zn contents were measured using flame atomic absorption spectrophotometer.

Histochemical Detection of Hydrogen Peroxide and Superoxide

The $\text{HO}_2^{\bullet-}$ and $\text{O}_2^{\bullet-}$ were localized histochemically (Chen et al., 2010) by staining leaves with 1% 3,3-diaminobenzidine (DAB) and 0.1% nitroblue tetrazolium chloride (NBT) solution, respectively.

Lipid Peroxidation, H_2O_2 Content and $\text{O}_2^{\bullet-}$ Generation Rate

The level of lipid peroxidation was measured by estimating (MDA, a product of lipid peroxidation) using TBA according to Heath and Packer (1968) with modifications (Hasanuzzaman et al., 2011). Hydrogen peroxide (H_2O_2) was assayed according to Yu et al. (2003) by extracting leaves in potassium-phosphate buffer (pH 6.5) (centrifuging at $11,500 \times g$), then adding it to a mixture of TiCl_4 in 20% H_2SO_4 (v/v) and the resulting solution was measured spectrophotometrically at 410 nm. To measure rate of $\text{O}_2^{\bullet-}$ generation leaves were homogenized in K-P buffer solution, centrifuged at $5000 \times g$. Supernatant was mixed with extraction buffer and hydroxylamine hydrochloride for incubation; then mixed with sulfanilamide and naphthylamine, incubated at 25°C . The absorbance was measured at 530 nm (Nahar et al., 2016).

Extraction and Measurement of Ascorbate and Glutathione

Leaves (0.5 g) were homogenized in 5% meta-phosphoric acid containing 1 mM EDTA (centrifuged at $11,500 \times g$) for 15 min at 4°C and the supernatant was collected for analysis of ascorbate

and glutathione. Ascorbate content was determined following the method of Hasanuzzaman et al. (2011). To determine total ascorbate, the oxidized fraction was reduced by adding 0.1 M dithiothreitol for 1 h at room temperature and then read at 265 nm using 1.0 unit AO. Oxidized ascorbate (DHA) content was determined by subtracting reduced AsA from total AsA. The glutathione pool was assayed according to previously described methods (Yu et al., 2003) with modifications as described by Hasanuzzaman et al. (2011). Standard curves with known concentrations of GSH and GSSG were used. The content of GSH was calculated by subtracting GSSG from total GSH.

Protein Determination and Enzyme Extraction and Assays

The protein concentration was determined according to Bradford (1976). Leaves were homogenized with 50 mM K-P (potassium phosphate) buffer (pH 7.0) containing 100 mM KCl, 1 mM AsA, 5 mM β -mercaptoethanol, and 10% (w/v) glycerol, centrifuged at $11,500 \times g$ and supernatants were used for enzyme activity assay. SOD (EC 1.15.1.1) activity (El-Shabrawi et al., 2010), CAT (EC: 1.11.1.6) activity (Hasanuzzaman et al., 2011), APX (EC: 1.11.1.11) (Nakano and Asada, 1981), MDHAR (EC: 1.6.5.4) (Hossain et al., 1984), DHAR (EC: 1.8.5.1) (Nakano and Asada, 1981), GR (EC: 1.6.4.2) (Hasanuzzaman et al., 2011), and GST (EC: 2.5.1.18) (Hossain et al., 2006) activities were measured following standard methodologies. GPX (EC: 1.11.1.9), Gly I (EC: 4.4.1.5), Gly II (EC: 3.1.2.6) activity was measured following Hasanuzzaman et al. (2011). LOX (EC 1.13.11.12) activity was estimated as described in Nahar et al. (2016).

Methylglyoxal Level

Leaves were homogenized in 5% perchloric acid and centrifuged at $11,000 \times g$, supernatant was decolorized, neutralized, MG content was estimation by adding sodium dihydrogen phosphate and N-acetyl-L-cysteine to a final volume of 1 mL. Formation of the product N- α -acetyl-S-(1-hydroxy-2-oxo-prop-1-yl)cysteine was recorded after 10 min at a wavelength of 288 nm (Wild et al., 2012).

Measurement of Free Polyamine Content

Endogenous free PAs were estimated according to Kotzabasis et al. (1993). In brief, leaf tissue (0.1 g) was homogenized in 1 mL of 5% (v/v) cold perchloric acid (PCA). The homogenates were kept at 2°C for 2 h and centrifuged at $15,000 \times g$ for 20 min. The supernatant was collected and stored at 2°C . PA standards, free PAs were benzoylated. Briefly, 1 ml of 2N NaOH and 10 μl of benzoylchloride were added to 200 μl of the PA aliquots and vortexed for 30 s. After 20 min incubation at 25°C , 2 ml of saturated NaCl was added to stop the reaction. The benzoyl – PAs were gently mixed with 2 ml diethyl ether. After properly vortex, upper portion was taken and the ether phase was collected and evaporated to dryness in a water bath (60°C). The benzoyl-PAs were redissolved in 200 μl of 64% (v/v) methanol and analyzed by HPLC. A 20 μl of the benzoylated extract was injected into the C18 reverse phase HPLC column (4.6 mm \times 100 mm, 5 μm particle size,

INTRODUCTION

Plants show susceptibility to various stresses under both wild and cultivated conditions. In many areas of the world, high concentrations of salt in the soil are commonly encountered phenomenon. Salinity is a complex stress that involves both ionic and osmotic components. Salt stress adversely affects plant development and productivity by generating ion toxicity, inducing nutritional deficiencies, creating osmotic stress and water deficits, and decreasing photosynthesis (Shabala and Munns, 2012).

Reactive oxygen species are partly reduced forms of oxygen and by-products of aerobic metabolism and stronger oxidants than molecular oxygen. ROS are overproduced under stress condition and can potentially damage the intracellular machinery. ROS may include free radicals such as superoxide anion ($O_2^{\bullet-}$) and hydroxyl radical ($\bullet OH$), as well as non-radical molecules like hydrogen peroxide (H_2O_2) and singlet oxygen (1O_2). Plants can scavenge a certain amount of ROS generated as by-products of aerobic metabolism, but under stress conditions, including salt stress, the resulting photoinhibition often creates an imbalance between ROS generation and scavenging due to generation of excessive amount of ROS, thereby resulting in oxidative stress, causing peroxidation of lipids, oxidation of proteins, inhibition of enzyme activity, injury to nucleic acids, activation of programmed cell death (PCD) pathways, and eventually leading to death of the cells (Gill and Tuteja, 2010; Hasanuzzaman et al., 2013). The antioxidant defense machinery is a ROS scavenging system, which protects plants from oxidative damage. Efficient enzymatic (superoxide dismutase, SOD; catalase, CAT; ascorbate peroxidase, APX; glutathione reductase, GR; monodehydroascorbate reductase, MDHAR; dehydroascorbate reductase, DHAR; glutathione peroxidase, GPX; guaiacol peroxidase, GOPX; and glutathione S-transferase, GST) and non-enzymatic (ascorbate, AsA; glutathione, GSH; phenolic compounds, alkaloids, non-protein amino acids, and α -tocopherols) components form an antioxidant defense system to protect overproduction of ROS and to prevent plant cells from experiencing oxidative damage (Gill and Tuteja, 2010).

Methylglyoxal (MG) is a transition-state intermediate product of the triose-phosphates of the glycolysis pathway in eukaryotic cells. High accumulation of MG is toxic and inhibits cell proliferation, causes degradation of proteins by modifying arginine, lysine, and cysteine residues, forms adducts with guanyl nucleotide in DNA, and inactivates antioxidant defense system, and causes oxidative stress (Yadav et al., 2005). The glyoxalase system comprises of two enzymes, gly I (glyoxalase I) and gly II (glyoxalase II), which catalyze the detoxification reaction of MG to D-lactate (Yadav et al., 2005). Although the glyoxalase enzymes have been extensively studied in microbial and animal systems (Thornalley, 1990), the biological significance of this pathway in plants and under stress condition is just beginning to be explored (Yadav et al., 2005).

The polyamines (PAs) are low-molecular-weight organic cations found in a wide range of organisms, where they perform diverse biological functions (Kusano et al., 2008). Putrescine (Put), spermidine (Spd), and spermine (Spm) are the most

common PAs. The levels of PAs frequently increase during stress and they play roles in enhancing plant stress tolerance. Acting as molecular chaperones, PAs bind to negatively charged surfaces and protect membranes and biomolecules (Kusano et al., 2008). PAs may also act as ROS and free-radical scavengers and activate the antioxidant enzyme machinery which help to reduce oxidative stress and subsequent membrane injury and electrolyte leakage (Gill and Tuteja, 2010; Kubiś et al., 2014). PAs interact with some signaling molecules (Kusano et al., 2008). PAs have been confirmed to be unique polycationic metabolites involved in the direct blockage of a variety of cation and K^+ -selective channels, vacuolar-type channels, and ammonium channels (Dobrovinskaya et al., 1999). PAs restrict plasma membrane Na^+ influx and $NaCl$ -induced K^+ efflux and shoot-to-root K^+ recirculation (Shabala et al., 2007; Zhao et al., 2007). Despite their range of protective effects under stress condition, acute application of PAs can cause endogenous PAs catabolism in the apoplast, which is responsible for ROS-induced oxidative damage (Di Tomaso et al., 1989). The diversity of the physiological functions of PAs has made understanding the detailed mechanisms of the protective function of PAs in response to stress particularly challenging.

Salt stress tolerance is a polygenic trait. Some of the major traits affected by salinity include the maintenance of membrane transport activity and ionic homeostasis, compartmentalization of ions at the cellular and whole-plant level (which may include Na^+ exclusion from root uptake, intra-cellular Na^+ sequestration, cytosolic K^+ retention), osmotic adjustment through synthesis of compatible solutes, oxidative stress tolerance, biomembrane protection, and MG toxicity tolerance (Yadav et al., 2005; Shabala and Munns, 2012; Rahman et al., 2016). Changes in ranges of biochemical pathways and the induction of hormones and signaling molecules activate stress responsive genes that impart salt stress tolerance (Adem et al., 2014). Polyamines have diverse mechanism to enhance salt tolerance. Under salt stress PAs proved to be efficient in modulating membrane integrity and ion transport processes, vacuolar ion channels, shoot, and root growth (Di Tomaso et al., 1989; Dobrovinskaya et al., 1999), photosynthesis (Kotzabasis et al., 1993; Duan et al., 2008; Shu et al., 2012), antioxidative responses (Sheekand et al., 2008; Iqbal et al., 2014); Na^+ and nutrient influx/efflux (Shabala et al., 2007; Pottosin et al., 2012). However, the actual mechanism by which PAs modulate these physiological processes is not yet known. A few components of antioxidant system were studied in response to PAs modulation. Polyamine-induced MG detoxification process is unknown in plants. In the present study, we have investigated the roles of PAs in maintaining ionic and nutrient homeostasis, inducing osmotic adjustment, and protecting mung bean plant cells from oxidative damage and MG toxicity under salt stress. The results presented here provide new insight into the role of PAs in enhancing salt tolerance and mitigating salt stress-induced damage in mung bean seedlings by the coordinated action of PAs on ion homeostasis, antioxidant defense, and the glyoxalase system.



Polyamines Confer Salt Tolerance in Mung Bean (*Vigna radiata* L.) by Reducing Sodium Uptake, Improving Nutrient Homeostasis, Antioxidant Defense, and Methylglyoxal Detoxification Systems

Kamrun Nahar^{1,2}, Mirza Hasanuzzaman^{3*}, Anisur Rahman^{1,3}, Md. Mahabub Alam¹, Jubayer-Al Mahmud^{1,4}, Toshisada Suzuki⁵ and Masayuki Fujita¹

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica – Universidade Nova
de Lisboa, Portugal

Reviewed by:

Li-Song Chen,
Fujian Agriculture and Forestry
University, China
Miguel López Gómez,
University of Granada, Spain

*Correspondence:

Mirza Hasanuzzaman
mhzsauag@yahoo.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 09 May 2016

Accepted: 12 July 2016

Published: 28 July 2016

Citation:

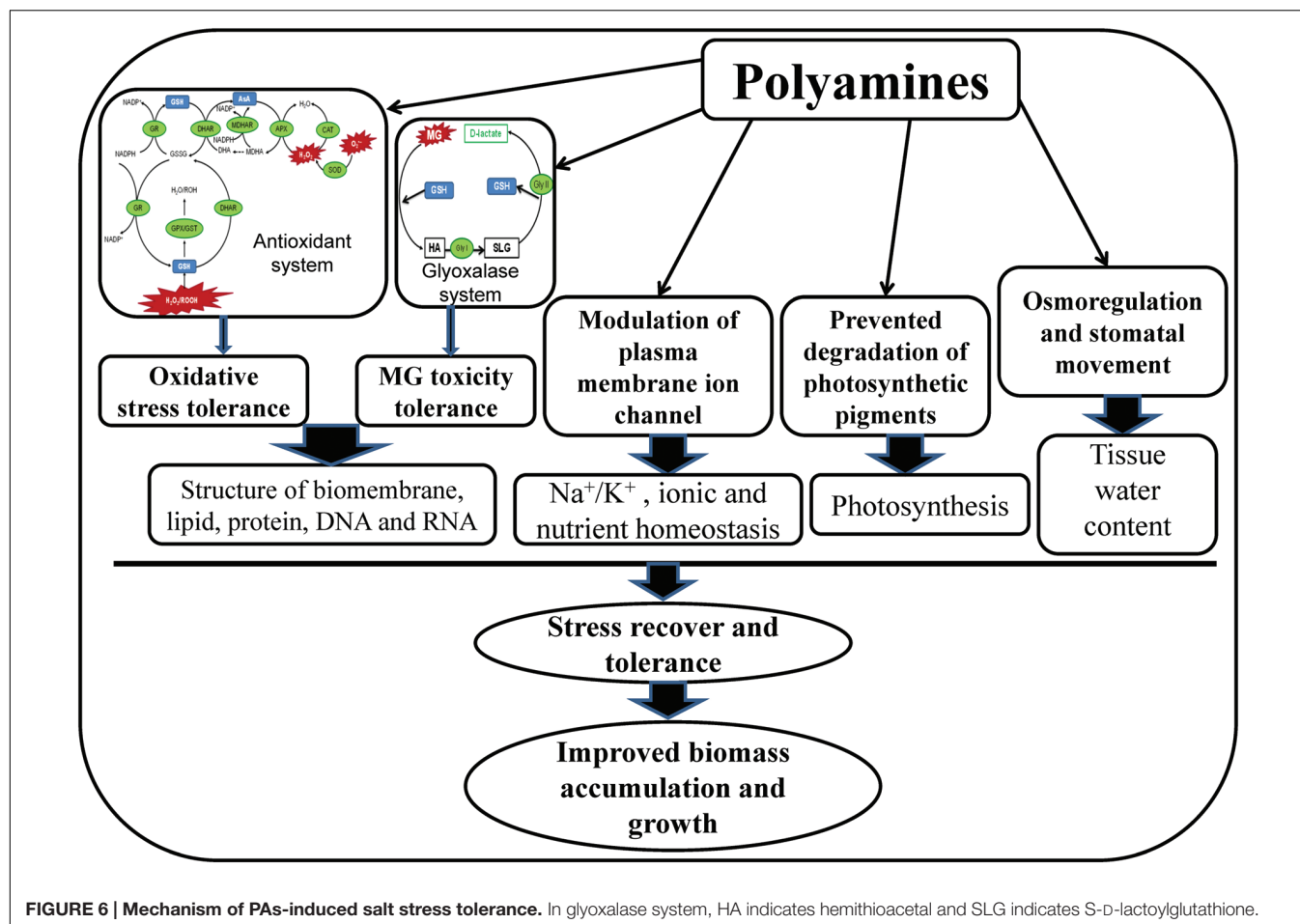
Nahar K, Hasanuzzaman M,
Rahman A, Alam MM, Mahmud J-A,
Suzuki T and Fujita M (2016)
Polyamines Confer Salt Tolerance
in Mung Bean (*Vigna radiata* L.) by
Reducing Sodium Uptake, Improving
Nutrient Homeostasis, Antioxidant
Defense, and Methylglyoxal
Detoxification Systems.
Front. Plant Sci. 7:1104.
doi: 10.3389/fpls.2016.01104

¹ Laboratory of Plant Stress Responses, Faculty of Agriculture, Kagawa University, Kagawa, Japan, ² Department of Agricultural Botany, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, ³ Department of Agronomy, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, ⁴ Department of Agroforestry and Environmental Science, Faculty of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, ⁵ Biomass Chemistry Laboratory, Bioresource Science for Manufacturing, Department of Applied Bioresource Science, Faculty of Agriculture, Kagawa University, Kagawa, Japan

The physiological roles of PAs (putrescine, spermidine, and spermine) were investigated for their ability to confer salt tolerance (200 mM NaCl, 48 h) in mung bean seedlings (*Vigna radiata* L. cv. BARI Mung-2). Salt stress resulted in Na toxicity, decreased K, Ca, Mg, and Zn contents in roots and shoots, and disrupted antioxidant defense system which caused oxidative damage as indicated by increased lipid peroxidation, H₂O₂ content, O₂^{•−} generation rate, and lipoxxygenase activity. Salinity-induced methylglyoxal (MG) toxicity was also clearly evident. Salinity decreased leaf chlorophyll (chl) and relative water content (RWC). Supplementation of salt affected seedlings with exogenous PAs enhanced the contents of glutathione and ascorbate, increased activities of antioxidant enzymes (dehydroascorbate reductase, glutathione reductase, catalase, and glutathione peroxidase) and glyoxalase enzyme (glyoxalase II), which reduced salt-induced oxidative stress and MG toxicity, respectively. Exogenous PAs reduced cellular Na content and maintained nutrient homeostasis and modulated endogenous PAs levels in salt affected mung bean seedlings. The overall salt tolerance was reflected through improved tissue water and chl content, and better seedling growth.

Keywords: abiotic stress, salinity, polyamine, methylglyoxal, oxidative damage, ROS signaling

Abbreviations: AO, ascorbate oxidase; APX, ascorbate peroxidase; AsA, ascorbic acid (ascorbate); BSA, bovine serum albumin; CAT, catalase; CDNB, 1-chloro-2,4-dinitrobenzene; chl, chlorophyll; DHA, dehydroascorbate; DHAR, dehydroascorbate reductase; DTNB, 5,5'-dithio-bis (2-nitrobenzoic acid); EDTA, ethylenediaminetetraacetic acid; Gly I, glyoxalase I; Gly II, glyoxalase II; GR, glutathione reductase; GSH, reduced glutathione; GSSG, oxidized glutathione; GPX, glutathione peroxidase; GST, glutathione S-transferase; LOX, lipoxxygenase; MDA, malondialdehyde; MDHA, monodehydroascorbate; MDHAR, monodehydroascorbate reductase; MG, methylglyoxal; NADPH, nicotinamide adenine dinucleotide phosphate; NTB, 2-nitro-5-thiobenzoic acid; Pro, Proline; ROS, reactive oxygen species; RWC, relative water content; SLG, S-D-lactoylglutathione; SOD, superoxide dismutase; TBA, thiobarbituric acid; TCA, trichloroacetic acid.



degradation pathways, and regulation of the enzymes related to those pathways. Supplementation of PAs relaxed salt stress effects and prevented chl degradation which restored and increased the contents of leaf chl (Li et al., 2014).

Inhibition of photosynthesis results in an over-reduction of the photosynthetic electron transport chain reactions, and it redirects photon energy into processes that promote the production of ROS and oxidative damage under salt stress (Sheokand et al., 2008; El-Shabrawi et al., 2010; Gill and Tuteja, 2010; Shabala and Munns, 2012). Salt stress creates ionic toxicity which inactivates the vital enzymes involved in major physiological processes, disrupted photosystem activity and mitochondrial electron transport chain causing imbalance between production and scavenging ROS, thereby amplify oxidative stress. Salt induced physiological drought, osmotic stress and stomatal closure are also responsible for ROS overproduction (Shabala and Munns, 2012). Mung bean plants exposed to salt stress showed higher accumulations of ROS, including H_2O_2 and $\text{O}_2^{\bullet-}$, and showed higher LOX activity, which resulted in cellular oxidative damage, as reflected in increases in membrane lipid peroxidation or MDA levels (Figures 2C–F). Visual identification of H_2O_2 and $\text{O}_2^{\bullet-}$, detected by histochemical staining (with DAB and NBT, respectively), also reflected a similar pattern of ROS generation under salt

stress (Figures 2A,B). Exogenous application of PAs reduced LOX activity, contents and spots and of H_2O_2 and $\text{O}_2^{\bullet-}$, and lipid peroxidation in mung bean seedlings (Figures 2A–F). In their study, Ikbal et al. (2014) reported that PAs biosynthesis was correlated with reduction spots of H_2O_2 and $\text{O}_2^{\bullet-}$ in grapevine leaves. Reductions in LOX activity and reduced MDA contents under salt stress were also observed in previous studies (Nahar et al., 2015). PAs act as ROS and free-radical scavengers (Gill and Tuteja, 2010), enhance antioxidant machinery (Gill and Tuteja, 2010; Kubiś et al., 2014), PAs act as molecular chaperones and PAs bind to negatively charged surfaces and protect membranes and biomolecules (Kusano et al., 2008). Exogenous Put had a protective effect on salt-induced membrane damage in chick pea, where increased activities of antioxidant enzymes decreased ROS generation (Sheokand et al., 2008). Similar results were reported in grapevine plants where PAs biosynthesis increased significantly (Ikbal et al., 2014).

A balanced state of cellular ROS equilibrium is maintained by the interaction between enzymatic and non-enzymatic antioxidants which combat oxidative damage. Constitutes the first line enzymatic defense against ROS, SOD converts $\text{O}_2^{\bullet-}$ to H_2O_2 , and H_2O_2 is readily detoxified by CAT by through conversion of H_2O_2 to H_2O and O_2 (Gill and Tuteja, 2010). Mung bean seedlings showed increased SOD and decreased CAT

activity, acutely increased $O_2^{\bullet-}$ and H_2O_2 contents in salt-stressed seedlings, supported by previous studies (Nahar et al., 2016). PAs maintained same SOD and increased CAT activities (**Table 1**) in salt-stressed plants which overwhelmed overproduced $O_2^{\bullet-}$ and H_2O_2 levels and improved leaf appearance by reducing damage spots (of $O_2^{\bullet-}$ and H_2O_2) (**Figures 2A,B; Table 1**). APX, MDHAR, DHAR, and GR are the enzymes of the ascorbate-glutathione cycle which has major roles in ROS scavenging process and regenerating AsA and GSH. APX decomposes H_2O_2 via oxidation of AsA to DHA. Ascorbate reacts with $O_2^{\bullet-}$, H_2O_2 to form MDHA or DHA. This reaction leads to DHA accumulation, which is harmful for plant cells. AsA is regenerated from DHA by MDHAR and DHAR, where NADPH and GSH are used as electron donors, respectively (Gill and Tuteja, 2010). Increased APX activity (**Table 1**) and DHA content (**Figure 3B**) and decreased AsA content (**Figure 3A**) decreased the AsA/DHA ratio (**Figure 3C**) of mung bean seedlings corroborating higher content of ROS. The increase in GSH level was not significant in salt-affected mung bean seedlings in the present study (compared to control), which showed increased ROS content and lipid peroxidation (**Figure 2**). Application of PAs increased activity of DHAR (**Table 1**), contents of AsA and AsA/DHA ratio, decreased DHA content (**Figures 3A–C**), increased the GR activity (enzyme involved in recycling of GSH) (**Table 1**), contents of GSH and GSH/GSSG ratio, decreased GSSG content (**Figures 3D–F**) which played major roles in the ROS (**Figures 2D–F**) detoxification process (Gill and Tuteja, 2010). Previous studies revealed the regulatory function of PAs in enhancing AsA-GSH cycle components and other antioxidant components that contributed to stress tolerance. Exogenous Spd application to salinized nutrient solution increased the antioxidant enzyme activities, including SOD, peroxidase (POD), and CAT, which alleviated salt-induced membrane damage and photosynthesis inhibition and promoted an increase in PAs content (Duan et al., 2008). Polyamine biosynthesis was correlated with enhanced APX, POD, SOD, and MDHAR activities, increased AsA content, decreased DHA levels, increased AsA/DHA ratio, increased GSH levels, decreased GSSG levels, and increased GSH/GSSG ratios in grapevine leaf tissues. Therefore, enhancement of PAs biosynthesis contributed to salt stress tolerance by upregulating ROS scavenging action and protecting the photosynthetic apparatus from oxidative damage (Ikbal et al., 2014). Using GSH as a substrate, GPX catalyzes the reduction of H_2O_2 and organic lipid hydroperoxides (Gill and Tuteja, 2010). GST conjugates GSH to a range of reactive aldehydes and xenobiotics to convert water-soluble and less toxic products. GSTs bear peroxidase activity, which reduces oxidative stress (Gill and Tuteja, 2010). The mung bean seedlings in this experiment showed a higher GPX and GST activity under salt stress, when compared to unstressed controls. GPX activity increased but GST activity did not increase further after PAs addition to salt-stressed plants, when compared to salt stress alone (**Table 1**). PAs have protective effects on proteins/antioxidant enzymes; PAs form complexes with SOD, GPX, and CAT for which these enzymes function more efficiently, compared with isolated enzymes (Li et al., 2014).

Methylglyoxal is a highly cytotoxic compound that causes protein and DNA degradation. MG disrupts the antioxidant machinery and act as mediator for $O_2^{\bullet-}$ generation, causing an oxidative burst (Yadav et al., 2005). The mung bean seedlings had high MG content under salt stress. The glyoxalase system comprises of two enzymes, Gly I, and Gly II, which catalyze the detoxification reaction of MG to D-lactate. This detoxification process occurs mainly via two steps: Gly I converts MG to SLG utilizing GSH, while Gly II converts SLG to D-lactic acid, and regenerates GSH (Yadav et al., 2005). Exogenous Put, Spd, and Spm application enhanced Gly II activity and increased the contents of GSH, which reduced MG toxicity by reducing its content (**Figures 4A–C and 5D**). Glutathione homeostasis and enhanced glyoxalase system activity were the biomarkers for salt tolerance in a Pokkali cultivar of rice (El-Shabrawi et al., 2010). Hasanuzzaman et al. (2011) reported that enhanced salt tolerance in wheat seedlings was partly contributed by increased activities of Gly I and Gly II. Modulation of PAs and MG in polyethylene glycol-affected white spruce was reported by Kong et al. (1998). The glyoxalase system was studied mostly in animals and microbes but not extensively in plants. It can be considered in stress specific studies.

The ratio of free [(Spd + Spm)/Put] is considered more important than the individual contents of Put, Spd, and Spm. A high [(Spd + Spm)/Put] ratio is crucial for imparting plant stress tolerance (Duan et al., 2008; Yang et al., 2010). Decreases in free Spd and Spm and increases in free Put decreased the [(Spd + Spm)/Put] ratio (**Figures 5A–D**), which indicates increased susceptibility of mung bean plants to NaCl toxicity (Yang et al., 2010). Therefore, the increase in Spd and Spm contents and elevation of the [(Spd + Spm)/Put] free ratio by exogenous application PAs with salt stress (**Figure 5D**) was critical in improving NaCl tolerance, in agreement with other published results showing that modulation of PAs was correlated with an altered physiology and biochemistry toward development of plant stress tolerance (Duan et al., 2008; Yang et al., 2010).

Higher NaCl concentration in the soil primarily causes ionic and osmotic stress that interrupts water transportation and reduces stomatal conductance, the efficiency of photosystems, RuBisCo activity, CO_2 assimilation, and photosynthesis, thereby reducing growth in salt-affected plants (Shu et al., 2012). In the present study, the overall tolerance was estimated by seedling growth and development. Exogenous application of PAs to salt-affected mung bean seedlings resulted in better seedling growth, as determined by their higher plant height, root length, and DW, when compared to salt-affected seedlings (**Table 2**) which is supported by previous findings. Exogenous Spd positively affected the photosynthetic and xanthophyll cycle in salt affected cucumber seedlings. Spermidine alleviated the salt-mediated decline in photosynthetic efficiency by enhancing the maximum quantum efficiency and actual efficiency of photosystem II (PSII), and improving the net photosynthetic rate, which improved overall growth performance of cucumber seedlings (Shu et al., 2012). Polyamine biosynthesis was correlated with oxidative stress protection of photosynthetic apparatus, enhancement of PSII quantum yield in grapevine plants that improved photosynthesis and plant growth (Ikbal et al., 2014).

CONCLUSION

In present study, salt stress created ionic and osmotic stress, disrupted biochemical processes, and altered PAs metabolism, resulted phytotoxicity in mung bean seedlings. High cellular Na content, imbalances in the mineral nutrients, oxidative damage, MG toxicity, and growth inhibition were the characteristic symptoms of salt stress affected mung bean seedlings. Polyamines played diversified roles in imparting salt stress tolerance in mung bean seedlings in the present study. The possible mechanism of PAs-induced salt stress tolerance has been presented in **Figure 6**. Exogenous PAs supplementation to salt-stressed mung bean plants modulated the endogenous levels of Put, Spd, and Spm and increased the (Spd+Spm)/Put ratio which might have regulatory roles in alteration of physiological features and stress defense mechanism of salt affected mung bean seedlings. Exogenous PAs application in mung bean seedlings increased content of AsA and AsA/DHA ratio, GSH content and GSH/GSSG ratio, enhanced activities of CAT, DHAR, GR, GPX which reduced ROS production, oxidative stress and subsequence membrane lipid peroxidation. PAs application in salt affected seedlings increased the content of GSH and activities of Gly I and Gly II enzymes which played vital roles in reducing MG toxicity. Reduction of ROS and MG by PAs were manifested through biomembrane and biomolecules protection and reduced oxidative damage in mung bean seedlings. Application of PAs prevented Na influx/toxicity and improved nutrient homeostasis in salt affected mung bean seedlings probably by regulating

the plasma membrane ion channel. In salt affected mung bean seedlings, PAs increased Pro content which helped to increase leaf RWC, compared to salt stress alone. Osmoregulation and maintaining tissue water are vital for smooth running of physiological processes smoothly, stress recovery and tolerance which were imparted by PAs under salt stress. Preventing degradation, PAs improved photosynthetic pigment contents which might increase photosynthesis and growth of mung bean seedlings in present study.

AUTHOR CONTRIBUTIONS

KN, MH, and MF conceived and designed the experiments; KN, MA, AR, JM, and TS performed the experiments; MH analyzed the data; MF contributed reagents/materials/analysis tools; KN and MH wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

The first author is grateful to the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan for financial supports. We acknowledge Dr. Md. Motiar Rohman, Senior Scientific Officer, Bangladesh Agricultural Research Institute for his cordial cooperation and helping in measurement of enzymatic activities and other biochemical parameters.

REFERENCES

- Addinsoft (2015). *XLSTAT V. 2015.1.01: Data Analysis and Statistics Software for Microsoft Excel*. Paris: Addinsoft.
- Adem, G. D., Roy, S. J., Zhou, M., Bowman, J. P., and Shabala, S. (2014). Evaluating contribution of ionic, osmotic and oxidative stress components towards salinity tolerance in barley. *BMC Plant Biol.* 14:113. doi: 10.1186/1471-2229-14-113
- Arnon, D. T. (1949). Copper enzymes in isolated chloroplasts polyphenol oxidase in *Beta vulgaris*. *J. Plant Physiol.* 24, 1–15. doi: 10.1104/pp.24.1.1
- Barrs, H. D., and Weatherley, P. E. (1962). A re-examination of the relative turgidity technique for estimating water deficits in leaves. *Aust. J. Biol. Sci.* 15, 413–428. doi: 10.1071/B19620413
- Bates, L. S., Waldren, R. P., and Teari, D. (1973). Rapid determination of free proline for water stress studies. *Plant Soil* 39, 205–207. doi: 10.1007/BF00018060
- Bradford, M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding. *Anal. Biochem.* 72, 248–254. doi: 10.1016/0003-2697(76)90527-3
- Chen, F., Wang, F., Wu, F., Mao, W., Zhang, G., and Zhou, M. (2010). Modulation of exogenous glutathione in antioxidant defense system against Cd stress in the two barley genotypes differing in Cd tolerance. *Plant Physiol. Biochem.* 48, 663–672. doi: 10.1016/j.plaphy.2010.05.001
- Di Tomaso, J. M., Shaff, J. E., and Kochian, L. V. (1989). Putrescine-induced wounding and its effects on membrane integrity and ion transport processes in roots of intact corn seedlings. *J. Plant Physiol.* 90, 988–995. doi: 10.1104/pp.90.3.988
- Dobrovinskaya, O. R., Muñoz, J., and Pottosin, I. I. (1999). Inhibition of vacuolar ion channels by polyamines. *J. Membr. Biol.* 167, 127–140. doi: 10.1007/s002329900477
- Duan, J., Li, J., Guo, S., and Kang, Y. (2008). Exogenous spermidine affects polyamine metabolism in salinity-stressed *Cucumis sativus* roots and enhances short-term salinity tolerance. *J. Plant Physiol.* 165, 1620–1635. doi: 10.1016/j.jplph.2007.11.006
- El-Shabrawi, H., Kumar, B., Kaul, T., Reddy, M. K., Singla-Pareek, S. L., and Sopory, S. K. (2010). Redox homeostasis, antioxidant defense, and methylglyoxal detoxification as markers for salt tolerance in Pokkali rice. *Protoplasma* 245, 85–96. doi: 10.1007/s00709-010-0144-6
- Gill, S. S., and Tuteja, N. (2010). Polyamines and abiotic stress tolerance in plants. *Plant Signal. Behav.* 5, 26–33. doi: 10.4161/psb.5.1.10291
- Hasanuzzaman, M., Hossain, M. A., and Fujita, M. (2011). Nitric oxide modulates antioxidant defense and the methylglyoxal detoxification system and reduces salinity-induced damage of wheat seedlings. *Plant Biotechnol. Rep.* 5, 353–365. doi: 10.1007/s11816-011-0189-9
- Hasanuzzaman, M., Nahar, K., Alam, M. M., Roychowdhury, R., and Fujita, M. (2013). Physiological, biochemical, and molecular mechanisms of heat stress tolerance in plants. *Int. J. Mol. Sci.* 14, 9643–9684. doi: 10.3390/ijms14059643
- He, L. X., Nada, K., Kasukabe, Y., and Tachibana, S. (2002). Enhanced susceptibility of photosynthesis to low-temperature photoinhibition due to interruption of chill induced increase of S-adenosylmethionine decarboxylase activity in leaves of spinach (*Spinacia oleracea* L.). *Plant Cell Physiol.* 43, 196–206. doi: 10.1093/pcp/pcf021
- Heath, R. L., and Packer, L. (1968). Photoperoxidation in isolated chloroplast. I. Kinetics and stoichiometry of fatty acid peroxidation. *Arch. Biochem. Biophys.* 125, 189–198. doi: 10.1016/0003-9861(68)90654-1
- Hossain, M. A., Nakano, Y., and Asada, K. (1984). Monodehydroascorbate reductase in spinach chloroplasts and its participation in the regeneration of ascorbate for scavenging hydrogen peroxide. *Plant Cell Physiol.* 25, 385–395.
- Hossain, M. Z., Hossain, M. D., and Fujita, M. (2006). Induction of pumpkin glutathione S-transferase by different stresses and its possible mechanisms. *Biol. Plant.* 50, 210–218. doi: 10.1007/s10535-006-0009-1
- Ikbāl, F. E., Hernández, J. A., Barba-Espín, G., Koussa, T., Aziz, A., Faize, M., et al. (2014). Enhanced salt-induced antioxidative responses involve a contribution of polyamine biosynthesis in grapevine plants. *J. Plant Physiol.* 171, 779–788. doi: 10.1016/j.jplph.2014.02.006

- Kong, L., Attreel, S. M., and Fowke, L. C. (1998). Effects of polyethylene glycol and methylglyoxal bis(guanyldihydrazone) on endogenous polyamine levels and somatic embryo maturation in white spruce (*Picea glauca*). *Plant Sci.* 133, 211–220. doi: 10.1016/S0168-9452(98)00040-5
- Kotzabasis, K., Fotinou, C., Roubelakis-Angelakis, K. A., and Ghanotakis, D. (1993). Polyamines in the photosynthetic apparatus. Photosystem II highly resolved subcomplexes are enriched in spermine. *Photosynth. Res.* 38, 83–88. doi: 10.1007/BF00015064
- Kubiś, J., Floryszak-Wieczorek, J., and Arasimowicz-Jelonek, M. (2014). Polyamines induce adaptive responses in water deficit stressed cucumber roots. *J. Plant Res.* 127, 151–158. doi: 10.1007/s10265-013-0585-z
- Kusano, T., Berberich, T., Tateda, C., and Takahashi, Y. (2008). Polyamines: essential factors for growth and survival. *Planta* 228, 367–381. doi: 10.1007/s00425-008-0772-7
- Li, X., Gong, B., and Xu, K. (2014). Interaction of nitric oxide and polyamines involves antioxidants and physiological strategies against chilling-induced oxidative damage in *Zingiber officinale* Roscoe. *Sci. Hort.* 170, 237–248. doi: 10.1016/j.scienta.2014.03.026
- Nahar, K., Hasanuzzaman, M., Alam, M. M., and Fujita, M. (2015). Roles of exogenous glutathione in antioxidant defense system and methylglyoxal detoxification during salt stress in mung bean. *Biol. Plant.* 59, 745–756. doi: 10.1007/s10535-015-0542-x
- Nahar, K., Hasanuzzaman, M., Alam, M. M., Rahman, A., Suzuki, T., and Fujita, M. (2016). Polyamine and nitric oxide crosstalk: antagonistic effects on cadmium toxicity in mung bean plants through upregulating the metal detoxification, antioxidant defense, and methylglyoxal detoxification systems. *Ecotoxicol. Environ. Saf.* 126, 245–255. doi: 10.1016/j.ecoenv.2015.12.026
- Nakano, Y., and Asada, K. (1981). Hydrogen peroxide is scavenged by ascorbate-specific peroxidase in spinach chloroplasts. *Plant Cell Physiol.* 22, 867–880.
- Pottosin, I., Velarde-Buendía, A. M., Zepeda-Jazo, I., Dobrovinskaya, O., and Shabala, S. (2012). Synergism between polyamines and ROS in the induction of Ca²⁺ and K⁺ fluxes in roots. *Plant Signal. Behav.* 7, 1084–1087. doi: 10.4161/psb.21185
- Rahman, A., Nahar, K., Hasanuzzaman, M., and Fujita, M. (2016). Calcium supplementation improves Na⁺/K⁺ ratio, antioxidant defense and glyoxalase systems in salt-stressed rice seedlings. *Front. Plant Sci.* 7:609. doi: 10.3389/fpls.2016.00609
- Shabala, S., Cuin, T. A., and Pottosin, I. (2007). Polyamines prevent NaCl-induced K⁺ efflux from pea mesophyll by blocking non-selective cation channels. *FEBS Lett.* 581, 1993–1999. doi: 10.1016/j.febslet.2007.04.03
- Shabala, S., and Munns, R. (2012). “Salinity stress: physiological constraints and adaptive mechanisms,” in *Plant Stress Physiology*, ed. S. Shabala (Oxford: CAB International), 59–93. doi: 10.1079/9781845939953.0059
- Sheokand, S., Kumari, A., and Sawhney, V. (2008). Effect of nitric oxide and putrescine on antioxidative responses under NaCl stress in chickpea plants. *Physiol. Mol. Biol. Plants* 14, 355–362. doi: 10.1007/s12298-008-0034-y
- Shu, S., Yuan, L. Y., Guo, S. R., Sun, J., and Liu, C. J. (2012). Effects of exogenous spermidine on photosynthesis, xanthophyll cycle and endogenous polyamines in cucumber seedlings exposed to salinity. *Afr. J. Biotechnol.* 11, 6064–6074. doi: 10.5897/AJB11.1354
- Thornalley, P. J. (1990). The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. *Biochem. J.* 269, 1–11. doi: 10.1042/bj2690001
- Wild, R., Ooi, L., Srikanth, V., and Münch, G. (2012). A quick, convenient and economical method for the reliable determination of methylglyoxal in millimolar concentrations: the N-acetyl-L-cysteine assay. *Anal. Bioanal. Chem.* 403, 2577–2581. doi: 10.1007/s00216-012-6086-4
- Yadav, S. K., Singla-Pareek, S. L., Ray, M., Reddy, M. K., and Sopory, S. K. (2005). Transgenic tobacco plants overexpressing glyoxalase enzymes resist an increase in methylglyoxal and maintain higher reduced glutathione levels under salinity stress. *FEBS Lett.* 579, 6265–6271. doi: 10.1016/j.febslet.2005.10.006
- Yang, H., Shi, G., Wang, H., and Xu, Q. (2010). Involvement of polyamines in adaptation of *Potamogeton crispus* L. to cadmium stress. *Aquat. Toxicol.* 100, 282–288. doi: 10.1016/j.aquatox.2010.07.026
- Yu, C. W., Murphy, T. M., and Lin, C. H. (2003). Hydrogen peroxide-induces chilling tolerance in mung beans mediated through ABA-independent glutathione accumulation. *Funct. Plant Biol.* 30, 955–963. doi: 10.1071/FP03091
- Zhao, F., Song, C. P., He, J., and Zhu, H. (2007). Polyamines improve K⁺/Na⁺ homeostasis in barley seedlings by regulating root ion channel activities. *Plant Physiol.* 145, 1061–1072. doi: 10.1104/pp.107.105882

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Nahar, Hasanuzzaman, Rahman, Alam, Mahmud, Suzuki and Fujita. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Association Mapping for Fiber-Related Traits and Digestibility in Alfalfa (*Medicago sativa*)

Zan Wang^{1*}, Haiping Qiang¹, Haiming Zhao², Ruixuan Xu³, Zhengli Zhang¹, Hongwen Gao¹, Xuemin Wang¹, Guibo Liu² and Yingjun Zhang³

¹ Institute of Animal Sciences, Chinese Academy of Agriculture Sciences, Beijing, China, ² Institute of Dry Farming, Hebei Academy of Agriculture and Forestry Sciences, Hengshui, China, ³ Department of Grassland Science, China Agricultural University, Beijing, China

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Ezio Portis,
University of Torino, Italy
Hongwei Cai,
China Agricultural University, China

*Correspondence:

Zan Wang
wangzan@caas.cn

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 18 November 2015

Accepted: 04 March 2016

Published: 18 March 2016

Citation:

Wang Z, Qiang H, Zhao H, Xu R,
Zhang Z, Gao H, Wang X, Liu G and
Zhang Y (2016) Association Mapping
for Fiber-Related Traits and
Digestibility in Alfalfa (*Medicago*
sativa). *Front. Plant Sci.* 7:331.
doi: 10.3389/fpls.2016.00331

Association mapping is a powerful approach for exploring the molecular genetic basis of complex quantitative traits. An alfalfa (*Medicago sativa*) association panel comprised of 336 genotypes from 75 alfalfa accessions represented by four to eight genotypes for each accession. Each genotype was genotyped using 85 simple sequence repeat (SSR) markers and phenotyped for five fiber-related traits in four different environments. A model-based structure analysis was used to group all genotypes into two groups. Most of the genotypes have a low relative kinship (<0.3), suggesting population stratification not be an issue for association analysis. Generally, the Q + K model exhibited the best performance to eliminate the false associated positives. In total, 124 marker-trait associations were predicted ($p < 0.005$). Among these, eight associations were predicted in two environments repeatedly and 20 markers were predicted to be associated with multiple traits. These trait-associated markers will greatly help marker-assisted breeding programs to improve fiber-related quality traits in alfalfa.

Keywords: alfalfa, association mapping, fiber-related traits, Simple sequence repeat (SSR)

INTRODUCTION

Alfalfa (*Medicago sativa*) is one of the most important forage crops in the world due to its high biomass and choice nutritional profiles and it provides reliable sources of protein and minerals to animals. The main objective, however, in many alfalfa quality breeding programs is to improve the digestibility (Buxton and Redfearn, 1997) for poor stem digestibility would cause major loss in animal feeding values (Mowat et al., 1965). A research indicated minor improvement in alfalfa stem digestibility would impact agriculture economically (Jung and Allen, 1995). In these sense, efforts in traditional breeding of improving the quality traits as well as yield, resistance, and agronomic traits are necessary. The feeding quality traits are usually quantitative, i.e., controlled by multiple genes. Understanding the genetic architecture of these traits on the molecular level is of necessity for efficient molecular breeding.

Currently, linkage analysis via QTL mapping, genome-wide association mapping (GWAS), and joint-mapping by combining linkage and association analysis (Ed Buckler's NAM population in maize) are the three main methods for dissecting complex quantitative traits. Compared with the traditional linkage analysis based on mapping populations, association mapping, has been proposed as an alternative powerful tool to overcome limitations of pedigree based QTL mapping for it has higher mapping resolution, reduced research time, and greater allele number

(Zhu et al., 2008). By utilizing historical recombinations that break LDs (linkage disequilibrium), association mapping has been widely adopted for almost all major crop species for gene identification and QTL validation, as well as better understanding of the genetic basis of complex traits (Gupta et al., 2011; Jiang et al., 2014; Wei et al., 2014; Font i Forcada et al., 2015; Portis et al., 2015). Given the facts that the most forage plants have a short selection and breeding history, Li et al. (2011) performed a GWAS analysis to map the yield and stem composition in an alfalfa breeding population of 190 individuals based on 71 SSR markers, and identified only one SSR strongly associated with acid detergent fiber (ADF) and acid detergent lignin (ADL), respectively. In this study, the association analysis was conducted for five fiber-related traits using a panel of 336 alfalfa genotypes, partially derived from the alfalfa core collection set developed by Basigalup et al. (1995) plus some germplasm from China. The study aimed to identify desirable alleles which could show significant trait-marker associations for the improvement of digestibility in alfalfa.

MATERIALS AND METHODS

Plant Materials and Experimental Design

A total of 336 cultivated tetraploid alfalfa genotypes from 75 *M. sativa* subsp. *sativa* accessions were selected to construct the association mapping population (Table S1). Each accession was represented by four genotypes, except for the Chinese accessions represented by eight genotypes for each accession. Nine Chinese accessions were collected from National Herbage Germplasm Bank of China; two accessions from Syria, one from Libya, and one accession from Sudan provided by the Institute of Animal Science, Chinese Academy of Agricultural Science (Beijing, China); the rest 62 accessions were provided by USDA National Plant Germplasm System (NPGS). All genotypes were grown and clonally propagated. Field experiments were conducted at the experimental station at the Institute of Dry Farming, Hebei Academy of Agriculture, and Forestry Sciences (Hengshui, Hebei Province, 37°44'N; 115°42'E) in May 2012, and at the experimental station of Institute of Animal Science of CAAS (Changping, Beijing, 40°10'N; 116°13'E) in May 2014. The experiment at each location used a randomized completed block design with two replications, each of which contained six clones. Within each plot, each cloned plant was spaced by 30 cm and each row was spaced by 75 cm. The biomass above the ground were trimmed ~1 month after establishment and then regrew for the remainder of the season.

Phenotyping

Plants were tested under four environments (Hengshui, 2013, 2014, 2015; Changping, 2014, designed as 13HS, 14HS, 15HS, and 14CP, respectively). Plant leave tissue samples were ground and then passed through a 1-mm mesh screen (Cyclone Mill, UDY Mfg., Ft. Collins, CO). Aliquots of each sample were scanned by near-infrared reflectance spectroscopy. Measurements were obtained using a FOSS 5000 scanning monochromator (FOSS, Denmark) and recorded at 2-nm intervals between 1100 and 2500 nm. A subset of 100 samples was selected for the

calibration of spectroscopy using chemical analyses. Coefficients of determination (R^2) were 0.9317 for ADF, 0.9634 for neutral detergent fiber (NDF), 0.6769 for ADL, 0.4950 for NDF digestibility (*in vitro* 30 h, NDFD 30 h), and 0.9001 for NDFD 48 h.

Genotyping

Eighty-five polymorphisms SSR primers (Eujayl et al., 2004; Robins et al., 2007) were used for genotyping. DNA extraction, PCR amplification, electrophoresis, and SSR genotyping analysis were conducted according to the methods described by Qiang et al. (2015).

Data Analyses

Analysis of variance (ANOVA) of all phenotypic data based on the means of traits of each accession under four environments was conducted as model: Phen = genotypes + environments + e, where Phen as the phenotypic observation, genotypes as the genetic effect, environments as the effect of the four environments, and e as the residual. All analyses were performed using SAS 8.02 (SAS Institute, 1999). Broad-sense heritability (H^2) was calculated as the genotypic variance divided by the total variance.

A kinship matrix was calculated using SPAGeDi software (Hardy and Vekemans, 2002).

The association between the phenotypes and markers was performed using Tassel v2.1 software (Bradbury et al., 2007). Three models were tested, namely the simple general linear model (GLM, Naive-model), the structured association model (GLM, Q model), and the mixed linear model (MLM, Q + K model) (Yu et al., 2006). The marker-trait association was considered as significant using a threshold of $P < 0.005$.

RESULTS

Phenotypic Variation

The descriptive parameters of the five measured traits under four environments were shown in Table 1. In summary, the ADF ranged from 20.59 to 42.85%, with an average of 31.22–33.63%; the NDF ranged from 25.08 to 53.06% with a mean of 36.72–42.89%; the NDFD 30 h alternated from 12.28 to 26.75% with an average of 16.68–23.81%; the NDFD 48 h ranged from 11.02 to 22.4% with an average of 15.47–17.55%; the ADL varied from 2.70 to 8.66% with an average of 4.26–6.56%. All the datasets showed a normal or nearly normal distribution (Figure S1). The broad-sense heritability of most traits was relatively high, ranging from 63% for NDF 30 h to 76% for NDF 48 h (Table 1), except for the heritability of ADL (45.1%), indicating the majority of studied traits were dominated by the genetic factors rather than the environmental variations. All five traits were significantly influenced by genotypes, environments and genotype × environments interactions (Table 1).

Population Structure and Relative Kinship

The genetic relationships among the genotypes were investigated using a model-based Bayesian clustering method on the 85 SSR marker genotyping data. Two populations were

TABLE 1 | Phenotypic variation for five traits in alfalfa in four environments.

Traits	E	Min	Max	Mean \pm SD	H^2	G	E	G \times E
ADF (DM %)	13HS	25.3	41.9	33.63 \pm 2.85	73.6	***	***	***
	14HS	21.3	42.9	33.00 \pm 3.32				
	15HS	20.6	40.1	31.80 \pm 2.85				
	14CP	22.8	39.0	31.22 \pm 2.69				
NDF (DM %)	13HS	31.5	53.1	42.89 \pm 3.60	74.6	***	***	***
	14HS	25.1	50.6	38.74 \pm 3.82				
	15HS	25.5	45.6	36.72 \pm 3.24				
	14CP	28.4	46.8	38.11 \pm 3.10				
NDFD 30 h (NDF %)	13HS	12.3	21.8	16.68 \pm 1.46	63	***	***	***
	14HS	20.4	26.8	23.81 \pm 0.98				
	15HS	18.8	24.3	21.96 \pm 1.04				
	14CP	19.6	25.1	22.62 \pm 1.07				
NDFD 48 h (NDF %)	13HS	11.0	20.2	16.11 \pm 1.46	76	***	***	***
	14HS	11.3	22.0	17.55 \pm 1.57				
	15HS	12.5	22.4	17.23 \pm 1.40				
	14CP	11.5	19.9	15.47 \pm 1.21				
ADL (DM %)	13HS	5.11	8.66	6.56 \pm 0.70	45.1	***	***	***
	14HS	2.70	5.64	4.46 \pm 0.47				
	15HS	3.27	6.10	5.00 \pm 0.38				
	14CP	2.88	5.44	4.26 \pm 0.44				

E, environment; Min, minimum; Max, maximum; SD, standard deviation; G, genotype; H^2 , Broad-sense heritability.

Abbreviations of traits and E are explained in Materials and Methods.

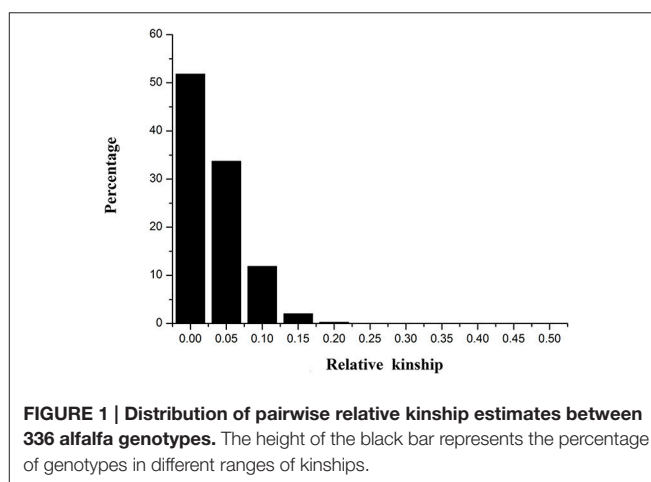
***Significant at $P < 0.001$.

identified by STRUCTURE software using a Bayesian approach, corresponding to China, and the rest of the world as indicated by Qiang et al. (2015). The kinship was estimated based on the 85 SSR data on 336 alfalfa genotypes. About 51.8% of the pairwise kinship estimates were equal to 0, while 99.7% of the relative kinship estimates were <0.2 in this alfalfa panel (Figure 1). These results indicated that most accessions have no or weak kinship with the other accessions in the panel, which might be due to the broad range collection of genotypes.

Association Analysis

For all five fiber-related traits, association analyses were conducted to assess the performance of three different models (Table 2 and Figure 2). Generally, the observed P -value from GLM greatly deviated from the expected P -value, followed by the Q model, while the P -value from the Q + K model was close to the expected P -value (Table 2 and Figure 2). The result indicated that the false positives were well controlled in the MLM model in the study. Therefore, subsequent analyses were done based on the MLM model.

Using the Q + K model, a total of 124 significant marker-trait associations was predicted under at least one environment (Table S2). For ADF trait associations, six, one, four, and 12 alleles were predicted as significant in 13HS, 14CP, 14HS, and



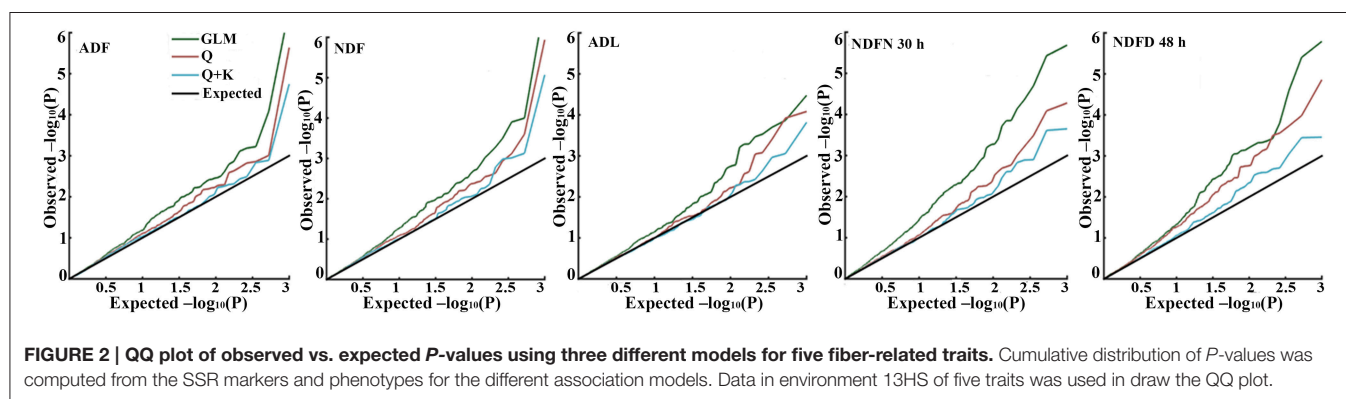
15HS data sets, respectively, with the explained phenotypic variance (R^2) ranging from 2.48 to 8.13% (Table S2). For ADL, five, one, seven, and six associated alleles were identified in four environments, respectively, with the R^2 varied from 2.56 to 9.66%. For NDF, six, one, four, and nine significant associated alleles were identified in four environments, respectively, with the R^2 from 2.76 to 8.28%. For NDF 30h, eight, two, six, and eight significant associated alleles were identified in four

TABLE 2 | Association summary for five fiber-related traits using three models in different environments.

Traits	E	GLM		Q model		Q + K model	
		No of alleles	R^2 (%)	No of alleles	R^2 (%)	No of alleles	R^2 (%)
ADF	13HS	16	2.5–8.4	7	2.8–6.8	6	2.5–6.0
	14HS	13	2.6–5.6	6	2.5–3.9	4	2.5–3.6
	15HS	16	2.6–5.2	12	2.5–3.6	12	2.5–3.6
	14CP	9	5.4–7.4	11	5.4–7.5	1	8.1
NDF	13HS	18	2.6–8.8	12	2.3–7.1	6	2.8–6.5
	14HS	14	2.7–5.5	10	5.5–7.7	1	8.28
	15HS	18	2.5–5.6	6	2.4–3.4	4	3.0–3.1
	14CP	9	5.5–7.4	10	2.4–3.7	9	2.9–3.8
NDFD 30h	13HS	32	2.5–7.0	12	2.1–4.4	8	2.5–4.1
	14HS	9	2.6–4.2	8	5.5–6.9	2	5.9–7.0
	15HS	24	2.5–5.5	7	2.6–3.9	6	2.7–3.6
	14CP	12	5.6–7.4	12	2.5–5.7	8	2.9–6.0
NDFD 48 h	13HS	37	2.5–7.1	23	2.3–5.6	11	2.5–4.1
	14HS	18	2.5–5.7	10	6.1–9.2	10	5.8–9.3
	15HS	25	2.6–7.3	6	2.3–4.6	4	2.9–4.9
	14CP	11	5.4–9.4	16	2.4–4.5	13	2.6–4.0
ADL	13HS	9	2.5–5.2	8	2.5–5.1	5	2.6–5.3
	14HS	17	2.6–5.5	13	5.2–9.2	1	9.7
	15HS	30	2.6–5.1	8	2.5–4.6	7	2.6–4.5
	14CP	8	5.5–7.8	16	2.6–5.2	6	3.8–4.8

Abbreviations of traits and E are explained in Materials and Methods.

R^2 , the explained phenotypic variance.



environments, respectively, with the R^2 from 2.52 to 6.98%. For NDF 48 h, eight, eight, three, and 10 significant associated alleles were detected in four environments, respectively, with the R^2 from 2.63 to 9.32%. Among these associated alleles, eight alleles were repeatedly observed in two environments (Table 3). For example, allele m13_173 associated with ADF was detected both in 14HS and 15HS. The allele m561-216 was associated with ADL both in 13HS and 14HS. In addition, among these associated alleles, 20 alleles were commonly associated with multiple fiber-related traits (Table S2). For example, the allele

m561_216 was associated with ADF, ADL, NDF, NDF30, and NDF 48 h.

The allele effect derived from significant marker-trait association was shown in Table S2. Among the markers associated with ADF, M115_183 had the most positive phenotypic effect (8.79), whereas m2_142 had the most negative phenotypic effect (−3.54). The alleles m215_182 and m2_142 had the most positive (9.95) and most negative (−4.11) phenotypic effect associated with NDF, respectively. For the NDFD 30 h, m190_205 had the most positive phenotypic

TABLE 3 | Summary of simple sequence repeat (SSR) alleles associated with fiber-related traits in at least two environments.

Trait	E	Associated alleles	Position	Chromosome	P-value	R ² (%)	Allele effects
ADF	14HS	m13_173	221	2	0.00377	2.54	3.32
	15HS	m13_173	221	2	0.00058	3.60	3.32
	14HS	m350_342	113	1	0.00329	2.72	1.47
	15HS	m350_342	113	1	0.00276	2.73	1.29
	13HS	m561_216	406	3	0.00001	5.95	−1.48
	14HS	m561_216	406	3	0.00073	3.60	−1.35
ADL	13HS	m561_216	406	3	0.00005	5.25	−0.36
	14HS	m561_216	406	3	0.00428	2.56	−0.16
NDF	14HS	m13_173	221	2	0.00131	3.13	4.26
	15HS	m13_173	221	2	0.00047	3.76	3.84
	14HS	m350_342	113	1	0.00186	3.05	1.81
	15HS	m350_342	113	1	0.00146	3.12	1.56
	13HS	m561_216	406	3	0.00001	6.46	−1.92
	14HS	m561_216	406	3	0.00170	3.10	−1.45
NDFD 30 h	14HS	m350_342	113	1	0.00318	2.80	0.45
	15HS	m350_342	113	1	0.00002	5.96	0.69
NDFD 48 h	14CP	m115_183	913	7	0.00145	7.87	3.97
	15HS	m115_183	913	7	0.00278	2.71	3.71
	14HS	m13_173	221	2	0.00014	4.40	2.07
	15HS	m13_173	221	2	0.00153	2.97	1.43
	14CP	m5_152	517	4	0.00186	7.04	3.87
	15HS	m5_152	517	4	0.00235	2.77	3.67
	14CP	m520_134	791	7	0.00171	8.62	3.91
	15HS	m520_134	791	7	0.00241	2.99	3.67
	14CP	m520_137	793	7	0.00171	8.62	3.91
	15HS	m520_137	793	7	0.00241	2.99	3.67
	14HS	m561_216	406	3	0.00056	3.72	−0.65
	13HS	m561_216	406	3	0.00035	4.06	−0.63
	14HS	m710_236	663	5	0.00231	2.90	0.96
	15HS	m710_236	663	5	0.00303	2.63	0.84

Abbreviations of traits and environments are explained in Materials and Methods.
R², the explained phenotypic variance.

effect (2.18), whereas m199_289 had the most negative phenotypic effect (−4.03). Among the alleles associated with NDFD 48 h, m225_203 had the most positive phenotypic effect (5.88), whereas m338_268 had the most negative phenotypic effect (−4.36). For the ADL, m53_131 had the most positive phenotypic effect (1.69), whereas m53_176 had the most negative phenotypic effect (−1.2). Also, m13, the individuals carrying the allele 170 bp had a lower ADL and NDFD 48 h than those carrying alleles 173 bp (Table S2). For m2, the individuals carrying the allele 136 bp had a lower NDFD48h than those carrying alleles 140 bp (Table S2). For m225, the individuals carrying the allele 191 bp had a lower NDFD48h than those carrying alleles 203 bp (Table S2). For m53, the individuals carrying the allele 176 bp had a lower ADL than those carrying alleles 131 bp (Table S2).

DISCUSSION

Association mapping has increasingly become a viable approach for the genetic dissection of quantitative traits. Due to the diverse geographical origins, the germplasm panel may contain either population structure or familial relatedness (Yu and Buckler, 2006). One of the limitations of association mapping studies is the easy detection of false positives associations caused by the existence of the genetic structure in the populations studied (Flint-Garcia et al., 2005). Several researches reported, in the structured association population, the mixed model (Q + K) showed a significant improvement in goodness of fit for traits (Flint-Garcia et al., 2005; Yu et al., 2006). In this panel, association analysis was conducted for five fiber-related traits in four environments using the GLM—simple model, the Q model, and the Q + K model. The alfalfa association populations used

in this study contained population structure but no obvious familial relationships (Figure 1). The quantile–quantile (QQ) plots indicated that the Q + K model performed best for all five fiber-traits. It seems that the Q + K model was sufficient to minimize false-positive associations, especially for some traits not influenced by population structure, which was consistent with other model simulations and comparisons (Yu et al., 2006; Zhu et al., 2008). All the results indicate that model testing for quantitative traits is necessary for increasing the accuracy of association.

There was no previous study on alfalfa fiber trait mapping or association using molecular markers. A total of 124 alleles from 38 markers accounted for phenotypic variation with 2.46–9.66% were identified as associated with five fiber-related traits based on association analysis using the Q + K model (Table S2). These associated alleles were not consistent with the previous studies of Li et al. (2011) which may be explained by the different markers and different population used in the two studies. These was also observed in previous studies on linkage mapping and association mapping which found that different mapping populations detected different QTL regions (Agrama et al., 2007; Zhang et al., 2014).

Most of the loci that were associated with the five traits could only be identified in a specific environment, indicating that the fiber-related traits in the study are variously influenced by the environment. However, some stable associations were identified in our study, such as the allele m350_342 which located in chromosome 1 were repeatedly detected in two environments and associated with ADF, NDF, NDFN30h, and NDFD48h. Three markers, m115_183, m520_134, and m520_137, located in chromosome 7 were repeatedly detected in two environments and associated with NDFD48h. Markers with significant traits associated over multiple environments may indicate that the associated genes are more stably expressed (i.e., less environmental influence) (Ray et al., 2015). A low threshold, $P < 0.005$, was used to detect the marker-trait association due to the limited number of marker used in this study. If high-density DNA polymorphism datasets are used for association mapping, additional markers with high $-\log(P\text{-value})$ may be obtained.

Among these associated alleles, different distributed patterns were observed among eight chromosomes in alfalfa. Eight alleles from seven markers which associated all five traits were observed in the chromosome 1, while only one allele of one marker which associated two traits was observed in the chromosome 6 (Table S2). In the study, 20 markers were associated with more than one traits indicated these traits were correlative each other. Interestingly, the markers, m225, and m338, reportedly

associated with yield (Li et al., 2011), was found associated with NDFD 30h and NDFD48h in this study, suggesting a correlation between these traits as assessed by the SSR or these traits are controlled by the same or neighboring regions in the genome. The explained phenotypic variance of all associated alleles ranged from 2.46 to 9.66%, with mean of 3.84%. The result indicated that the fiber-related traits were complex in nature, i.e., controlled by multiple genes without obvious major effects.

The present study is the first attempt in associating alfalfa fiber-related traits with the genotyping results derived from SSR markers using a diverse set of global collection of alfalfa genotypes. Our results demonstrated that this alfalfa panel is suitable for association mapping analysis targeting complex quantitative traits with optimal association models. The markers associated to the QTLs in the study can be effectively used in further alfalfa marker assisted breeding programmers for introgression of alleles into locally well adapted germplasm.

AUTHOR CONTRIBUTIONS

ZW designed the experiments performed the statistical analysis and drafted the manuscript. HQ performed SSR genotyping. HZ, GL, ZZ, RX, and YZ conducted the quality analysis. XW and HG revised manuscript. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by the National Basic Research Program of China (No. 2014CB138703), National Natural Science Foundation of China (No. 31272495), and Agricultural Science and Technology Innovation Program (No. ASTIP-IAS10) of China. We thank the anonymous reviewers for constructive comments on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00331>

Table S1 | Accession No. origin, improvement status, cultivar name, and No. of genotypes sampled of 336 alfalfa genotypes.

Table S2 | List of associated SSR alleles of five fiber-related traits in different environments.

Figure S1 | Histograms showing frequency distribution of five fiber-related traits in different environments in the study. The y-axis denotes the value of frequency, whereas the x-axis shows resultant groups of genotypes.

REFERENCES

- Agrama, H. A., Eizenga, G. C., and Yan, W. (2007). Association mapping of yield and its components in rice cultivars. *Mol. Breed.* 1, 341–356. doi: 10.1007/s11032-006-9066-6
- Basigalup, D. H., Barnes, D. K., and Stucker, R. E. (1995). Development of a core collection for perennial *Medicago* plant introductions. *Crop Sci.* 35, 1163–1168. doi: 10.2135/cropsci1995.0011183X003500040042x

- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Buxton, D. R., and Redfearn, D. D. (1997). Plant limitations to fiber-digestion and utilization. *J. Nutr.* 127, 814S–818S.
- Eujayl, I., Sledge, M. K., Wang, L., May, G. D., Chekhovskiy, K., Zwonitzer, J. C., et al. (2004). *Medicago truncatula* EST-SSRs reveal cross-species

- genetic markers for *Medicago* spp. *Theor. Appl. Genet.* 108, 414–422. doi: 10.1007/s00122-003-1450-6
- Flint-Garcia, S. A., Thuillet, A. C., Yu, J., Pressoir, G., Romero, S. M., Mitchell, S. E., et al. (2005). Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44, 1054–1064. doi: 10.1111/j.1365-3113.2005.02591.x
- Font i Forcada, C., Velasco, L., Socías i Company, R., and Fernández i Martí, Á. (2015). Association mapping for kernel phytosterol content in almond. *Front. Plant Sci.* 6:530. doi: 10.3389/fpls.2015.00530
- Gupta, S., Kumari, K., Muthamilarasan, M., Parida, S. K., and Prasad, M. (2011). Association mapping of dynamic developmental plant height in common wheat. *Planta* 234, 891–902. doi: 10.1007/s00425-011-1434-8
- Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: A versatile computer program to analyses spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- Jiang, H., Huang, L., Ren, X., Chen, Y., Zhou, X., Xia, Y., et al. (2014). Diversity characterization and association analysis of agronomic traits in a Chinese peanut (*Arachis hypogaea* L.) mini-core collection. *J. Integr. Plant Biol.* 56, 159–169. doi: 10.1111/jipb.12132
- Jung, H. G., and Allen, M. S. (1995). Characteristics of plant cell walls affecting intake and digestibility of forages by ruminants. *J. Anim. Sci.* 73, 2774–2790.
- Li, X. H., Wei, Y. L., Moore, K. J., Michaud, R., Viands, D. R., Hansen, J. L., et al. (2011). Association mapping of biomass yield and stem composition in a tetraploid alfalfa breeding population. *Plant Genome* 4, 24–35. doi: 10.3835/plantgenome2010.09.0022
- Mowat, D. N., Fulkerson, R. S., Tossell, W. E., and Winch, J. E. (1965). The *in vitro* digestibility and protein content of leaf and stem portions of forages. *Can. J. Plant Sci.* 45, 321–331. doi: 10.4141/cjps65-065
- Portis, E., Cericola, F., Barchi, L., Toppino, L., Acciarri, N., Pulcini, L., et al. (2015). Association mapping for fruit, plant and leaf morphology traits in eggplant. *PLoS ONE* 10:e0135200. doi: 10.1371/journal.pone.0135200
- Qiang, H., Chen, Z., Zhang, Z., Wang, X., Gao, H., and Wang, Z. (2015). Molecular diversity and population structure of a worldwide collection of cultivated tetraploid Alfalfa (*Medicago sativa* subsp. *sativa* L.) germplasm as revealed by microsatellite markers. *PLoS ONE* 10:e0124592. doi: 10.1371/journal.pone.0124592
- Ray, J. D., Dhanapal, A. P., Singh, S. K., Hoyos-Villegas, V., Smith, J. R., Purcell, L. C., et al. (2015). Genome-wide association study of ureide concentration in diverse maturity group IV 1 soybean [*Glycine max* (L.) Merr.] accessions. G3 5, 2391–2403. doi: 10.1534/g3.115.021774
- Robins, J. G., Luth, D., Campbell, T. A., Baughan, G. R., He, C., Viands, D. R., et al. (2007). Genetic mapping of biomass production in tetraploid alfalfa. *Crop Sci.* 47, 1–10. doi: 10.2135/cropsci2005.11.0401
- SAS Institute (1999). *SAS User's Guide, Version 8*. Cary, NC: SAS Institute.
- Wei, Z. Z., Zhang, G. Y., Du, Q. Z., Zhang, J. F., Li, B. L., and Zhang, D. Q. (2014). Genetic diversity and association mapping of seed vigor in rice (*Oryza sativa* L.). *Planta* 239, 1309–1319. doi: 10.1007/s00425-014-2060-z
- Yu, J., and Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17, 155–160. doi: 10.1016/j.copbio.2006.02.003
- Yu, J., Pressoir, G., Briggs, W. H., Vroh, B. I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhang, P., Liu, X., Tong, H., Lu, Y., and Li, J. (2014). Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. *PLoS ONE* 9:e111508. doi: 10.1371/journal.pone.0111508
- Zhu, C., Gore, M., Buckler, E., and Yu, J. (2008). Status and prospects of association mapping in plants. *Plant Genome* 1, 5–20. doi: 10.3835/plantgenome2008.02.0089

Conflict of Interest Statement: The reviewer HC declared a shared affiliation, though no other collaboration, with the authors RX and YZ to the handling Editor, who ensured that the process nevertheless met the standards of a fair and objective review.

The other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Wang, Qiang, Zhao, Xu, Zhang, Gao, Wang, Liu and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Quinolizidine Alkaloid Biosynthesis in Lupins and Prospects for Grain Quality Improvement

Karen M. Frick^{1,2,3*}, Lars G. Kamphuis^{1,3}, Kadambot H. M. Siddique³, Karam B. Singh^{1,3} and Rhonda C. Foley¹

¹ Commonwealth Scientific and Industrial Research Organisation Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation, Floreat, WA, Australia, ² School of Plant Biology, The University of Western Australia, Crawley, WA, Australia, ³ The UWA Institute of Agriculture, The University of Western Australia, Perth, WA, Australia

OPEN ACCESS

Edited by:

Diego Rubiales,
Spanish National Research Council,
Spain

Reviewed by:

Michael Wink,
Heidelberg University, Germany
Frédéric Marsolais,
Agriculture and Agri-Food Canada,
Canada
Giovanna Boschin,
University of Milan, Italy

*Correspondence:

Karen M. Frick
karen.frick@csiro.au

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 08 December 2016

Accepted: 16 January 2017

Published: 31 January 2017

Citation:

Frick KM, Kamphuis LG,
Siddique KHM, Singh KB and
Foley RC (2017) Quinolizidine Alkaloid
Biosynthesis in Lupins and Prospects
for Grain Quality Improvement.
Front. Plant Sci. 8:87.
doi: 10.3389/fpls.2017.00087

Quinolizidine alkaloids (QAs) are toxic secondary metabolites found within the genus *Lupinus*, some species of which are commercially important grain legume crops including *Lupinus angustifolius* (narrow-leafed lupin, NLL), *L. luteus* (yellow lupin), *L. albus* (white lupin), and *L. mutabilis* (pearl lupin), with NLL grain being the most largely produced of the four species in Australia and worldwide. While QAs offer the plants protection against insect pests, the accumulation of QAs in lupin grain complicates its use for food purposes as QA levels must remain below the industry threshold (0.02%), which is often exceeded. It is not well understood what factors cause grain QA levels to exceed this threshold. Much of the early work on QA biosynthesis began in the 1970–1980s, with many QA chemical structures well-characterized and lupin cell cultures and enzyme assays employed to identify some biosynthetic enzymes and pathway intermediates. More recently, two genes associated with these enzymes have been characterized, however, the QA biosynthetic pathway remains only partially elucidated. Here, we review the research accomplished thus far concerning QAs in lupin and consider some possibilities for further elucidation and manipulation of the QA pathway in lupin crops, drawing on examples from model alkaloid species. One breeding strategy for lupin is to produce plants with high QAs in vegetative tissues while low in the grain in order to confer insect resistance to plants while keeping grain QA levels within industry regulations. With the knowledge achieved on alkaloid biosynthesis in other plant species in recent years, and the recent development of genomic and transcriptomic resources for NLL, there is considerable scope to facilitate advances in our knowledge of QAs, leading to the production of improved lupin crops.

Keywords: grain improvement, grain legume, lupin, *Lupinus angustifolius*, plant secondary metabolism, pulse, quinolizidine alkaloids

Abbreviations: BIA, benzylisoquinoline alkaloid; ECT/EFT-LCT/LFT, *p*-coumaroyl-CoA/feruloyl-CoA: (+)-epilupinine/(–)-lupinine *O*-coumaroyl/feruloyltransferase; HMT/HLT, tigloyl-CoA:(–)-13 α -hydroxymultiflorine/(+)-13 α -hydroxylupanine *O*-tigloyltransferase; L/ODC, lysine/ornithine decarboxylase; MIA, monoterpenoid indole alkaloid; NLL, narrow-leafed lupin; QA, quinolizidine alkaloid.

INTRODUCTION

Quinolizidine alkaloids (QAs) are secondary metabolites that occur mostly within the family Leguminosae and they can occur in the genus *Lupinus*, as well as in *Baptisia*, *Thermopsis*, *Genista*, *Cytisus*, *Echinosophora*, and *Sophora* (Ohmiya et al., 1995). Whilst QAs offer the plants protection against insect pests (Wink, 1992; Berlandier, 1996; Wang et al., 2000; Philippi et al., 2015), they cause a concern for the human consumption of lupin grain and lupin-based foods as high levels confer a bitter taste and may result in acute anticholinergic toxicity, characterized by symptoms such as blurry vision, headache, weakness, and nausea (Daverio et al., 2013). The lethal dose of QAs in children is estimated to be 11–25 mg total alkaloids kg⁻¹ body weight, while no fatal poisonings have been reported in adults (Allen, 1998; Petterson et al., 1998).

Lupinus is a diverse genus, though only four species have been domesticated and are agriculturally significant: *L. angustifolius* (NLL), *L. albus* (white lupin), *L. luteus* (yellow lupin), and *L. mutabilis* (pearl lupin; Petterson et al., 1998). These species have been domesticated relatively recently (Cowling et al., 1998) and as a consequence of this, undesirable traits such as the accumulation of QAs remain. While the grain has been used traditionally as an animal feed, it has gained recognition as a health food; it is high in protein and fiber and possesses certain beneficial nutraceutical properties (Petterson et al., 1997; Duranti et al., 2008; Sweetingham et al., 2008). QAs complicate the use of the grain for higher-value food purposes as they must remain below the industry threshold of 0.02% in Australia and some European countries (Cowling et al., 1998; Boschini et al., 2008; Jansen et al., 2009). QA levels can vary considerably from year to year under field conditions, often exceeding this threshold (Cowling and Tarr, 2004). As such, an understanding of the QA biosynthetic pathway is essential in assisting lupin breeders and farmers to produce high-value crops consistently.

Quinolizidine alkaloid biosynthesis has been studied far less extensively than some economically important alkaloids in other plant species, for example nicotine in *Nicotiana*, MIAs in *Catharanthus roseus*, i.e., vinblastine and vincristine, and BIAs in *Coptis japonica* and *Papaver somniferum*, i.e., berberine and morphine, respectively, which represent model species for understanding alkaloid biosynthesis. The past couple of decades have resulted in the identification of many genes involved in alkaloid biosynthesis in these species including biosynthetic genes, transcription factors and transporters, and the identification of enzymes and pathway intermediates through the development of genomic, transcriptomic, proteomic, and metabolomic data sets (Dewey and Xie, 2013; Hagel and Facchini, 2013; Beaudoin and Facchini, 2014; Pan et al., 2016). In the case of QAs, while the chemistry has been well characterized with more than 170 structures identified (Wink, 1993), the QA biosynthetic pathway has only been partially elucidated and information on enzymes and genes involved in QA biosynthesis is limited. Here, we discuss what is currently known about QA biosynthesis in lupin, draw on examples from model alkaloid species, and suggest future directions and ways to improve QA biosynthesis in lupin to produce higher-value lupin crops.

QUINOLIZIDINE ALKALOIDS AND BIOSYNTHESIS

Quinolizidine alkaloids are so-called because of their quinolizidine ring structure and can be divided into major structural classes: lupanine, angustifoline, lupinine, sparteine, multiflorine, aphylline, anagryne and cytosine, though the latter two are usually absent in lupins and more commonly found in *Thermopsis*, *Sophora*, *Echinosophora*, and *Genista* (Wink, 1987a; Ohmiya et al., 1995; Boschini and Resta, 2013). Each lupin species has a characteristic alkaloid profile (Table 1). Usually, only the presence of major QAs are reported—defined as individual QAs with levels ≥1% of total QAs—although many other QAs have been detected at trace levels in each of the lupin species (Wink et al., 1995). Of the major QAs in lupin grain, three of the four domesticated lupins share lupanine and 13α-hydroxylupanine. Each cultivated lupin species also has unique major QAs such as isolupanine and angustifoline for NLL, albine and multiflorine for *L. albus*, and lupinine for *L. luteus* (Table 1). The indole alkaloid gramine is also a major component in bitter *L. luteus* grain and the piperidine alkaloid ammodendrine is found in major quantities in *L. mutabilis* and minor quantities in *L. luteus* and *L. albus* grain (Wink et al., 1995; de Cortes Sánchez et al., 2005; Adhikari et al., 2012). QAs vary in their toxicity and their deterrence against insect pests. Sparteine and lupanine appear to be the two most toxic QAs to humans and laboratory animals (Allen, 1998; Petterson et al., 1998), with lupanine having the greatest impact on aphid survival, followed by indole alkaloid gramine, sparteine, lupinine, and 13α-hydroxylupanine and

TABLE 1 | Major QAs (≥1%) identified in seed of lupin species.

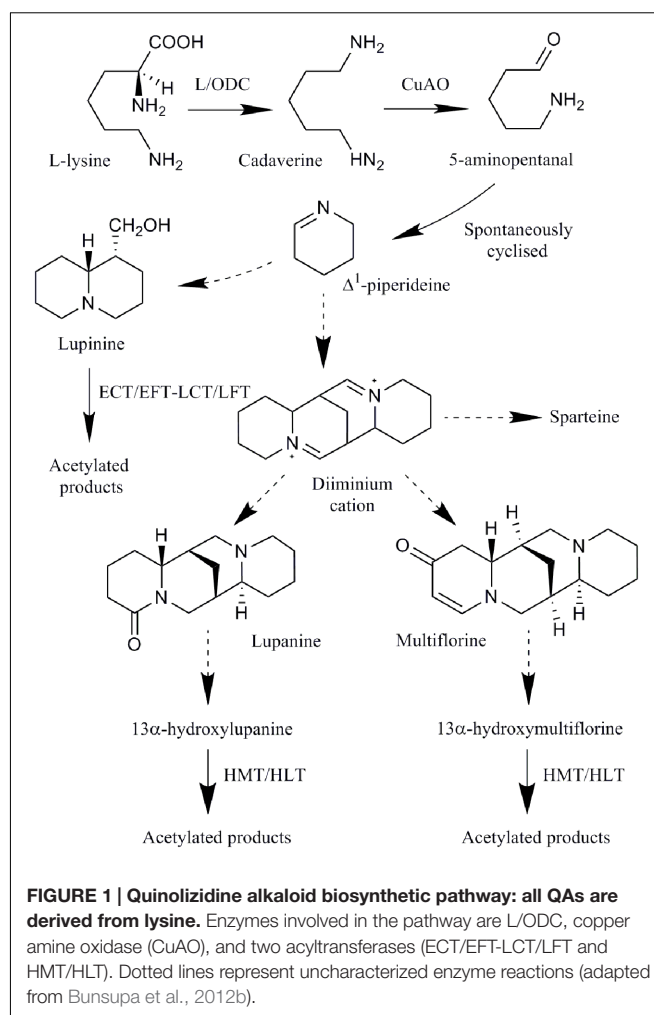
Lupin species	Major grain QAs (% of total alkaloids)	Reference
<i>L. angustifolius</i>	Lupanine (70%), 13α-hydroxylupanine (12%), angustifoline (10%) Lupanine (50.6%), 13α-hydroxylupanine (32.6%), angustifoline (10.4%), isolupanine (6.4%; average values)	Wink et al., 1995 Priddis, 1983
<i>L. albus</i>	Lupanine (70%), albine (15%), 13α-hydroxylupanine (8%), multiflorine (3%)	Wink et al., 1995
<i>L. luteus</i>	Lupinine (60%), sparteine (30%), unknown minor alkaloids (1%) Gramine* (10–89%), lupinine (3–60%), sparteine (1–20%), isoparteine (1–5%)	Wink et al., 1995 Adhikari et al., 2012
<i>L. mutabilis</i>	Lupanine (46%), sparteine (16%), 3β-hydroxylupanine (12%), 13α-hydroxylupanine (7%), ammodendrine* (2%), 13-angeloyloxylupanine (2%), tetrahydrohombifoline (2%), 11-12-dehydrosparteine (1%) angustifoline (1%), 13-tigoyloxylupanine (1%)	Wink et al., 1995

*Gramine and ammodendrine are indole and piperidine alkaloids, respectively.

angustifoline having the least impact (Ridsdill-Smith et al., 2004).

The biosynthesis of all QAs begins with the decarboxylation of L-lysine to form the intermediate cadaverine by a L/ODC such as the *Lupinus angustifolius* L/ODC (*La-L/ODC*; Leistner and Spenser, 1973; Bunsupa et al., 2012a) (**Figure 1**). Cadaverine then undergoes oxidative deamination, by a copper amine oxidase (CuAO), to yield 5-aminopentanal which is then spontaneously cyclized to Δ^1 -piperidine Schiff base (Leistner and Spenser, 1973; Golebiewski and Spenser, 1988; Bunsupa et al., 2012b). It has been suggested that in addition to these reactions, a series of reactions including Schiff base formations, aldol-type reactions, hydrolysis, oxidative deamination and coupling gives rise to the major structural QAs (e.g., lupanine and others; Dewick, 2002), with the diiminium cation proposed as an intermediate in the biosynthesis of tetracyclic alkaloids (e.g., lupanine, multiflorine, and sparteine; Fraser and Robins, 1984). These QAs can then be further modified by dehydrogenation, oxygenation, hydroxylation, glycosylation, or esterification to form a wide variety of structurally related QAs (Wink and Hartmann, 1982a; Saito et al., 1993; Ohmiya et al., 1995). The acyltransferase HMT/HLT forms acetylated products of 13 α -hydroxylupanine and 13 α -hydroxymultiflorine and a *L. albus* HMT/HLT (*LaHMT/HLT*) gene encoding this enzyme has been characterized (Saito et al., 1992; Suzuki et al., 1994; Okada et al., 2005). The acyltransferase ECT/EFT-LCT/LFT forms acetylated products of lupanine and epilupanine (Saito et al., 1992, 1993; Suzuki et al., 1994; Bunsupa et al., 2012b). *L. angustifolius* acyltransferase (*LaAT*) is suggested to be involved in the formation of QA esters, though its enzymatic function has not been confirmed (Bunsupa et al., 2011).

While only two genes have been identified in QA biosynthesis, the discovery of biosynthetic genes involved in the formation of other alkaloids may assist in identifying homologous genes in the QA pathway, for example, *La-L/ODC* was identified as a homolog of ODC, involved in the biosynthesis of a precursor for nicotine biosynthesis (Bunsupa et al., 2012a). Enzymes common in nicotine, MIA, BIA, as well as Amarylidiaceae alkaloid biosynthetic pathways include: methyltransferases, decarboxylases, oxidases, acyltransferases, cytochromes P450 (cP450s), oxidoreductases, demethylases, reductases, hydroxylases and coupling enzymes, and genes encoding many of these enzymes have been identified in *Nicotiana*, *C. roseus*, *C. japonica*, and *P. somniferum* (Bird et al., 2003; Dewey and Xie, 2013; Hagel and Facchini, 2013; Kilgore and Kutchan, 2016; Pan et al., 2016; Thamm et al., 2016). Many of these common types of enzymes are either known (i.e., decarboxylase, oxidase, and acyltransferases) or suggested (listed above) to play a role in QA biosynthesis. Transcriptome analysis has also identified several genes co-expressed with a putative *Sophora flavescens* L/ODC, encoding a major latex-like protein (MLP-like), a cP450, a ripening related protein and an uncharacterized protein (Han et al., 2015), which may also have roles in QA biosynthesis. MLP-like proteins may be involved in BIA biosynthesis, though their biological function is unknown, and the berberine bridge and berberine bridge-like enzymes catalyze oxidative reactions for the biosynthesis of BIAs



and *Nicotiana* alkaloids (Facchini et al., 1996b; Samanani et al., 2004; Kajikawa et al., 2011), possibly having similar roles in QA biosynthesis. Cytochromes P450 have a role in hydroxylation reactions, as well as other reactions, in MIA and BIA biosynthesis (Pauli and Kutchan, 1998; Thamm et al., 2016) and may be involved in QA hydroxylation reactions in the synthesis of derivatives of major structural QAs (**Figure 1**).

LOCALIZATION OF QA BIOSYNTHESIS AND TRANSPORT OF QAS

There is strong evidence for the synthesis of QAs in aerial tissues of lupin as opposed to roots: lupin L/ODC is localized to chloroplasts (Wink and Hartmann, 1982b; Bunsupa et al., 2012a), *La-L/ODC* transcript level is highest in young leaves of bitter NLL, while barely detectable in mature leaves, cotyledons, hypocotyls and roots (Bunsupa et al., 2012a), cadaverine is incorporated into lupanine in aerial tissue but not in roots (Wink, 1987b) and grafting experiments in lupin, whereby high-QA lupin scions are grafted onto low-QA lupin roots and vice versa, show that shoots are more important than roots in determining

overall plant QA content (Waller and Nowacki, 1978; Lee et al., 2007). The last step of lysine biosynthesis also takes place in the chloroplast (Mazelis et al., 1976; Wink and Hartmann, 1982b). Interestingly, *Lycopodium clavatum* L/ODC, with a role in biosynthesis of *Lycopodium* alkaloids which are also derived from lysine, is localized in the cytosol (Bunsupa et al., 2016) and perhaps the chloroplastic location of La-L/ODC increases its accessibility to lysine, rather than ornithine, as a substrate for the production of QAs.

The expression of *LaHMT/HLT* and *HMT/HLT* activity was associated with roots and hypocotyls of *Lupinus* plants (Saito et al., 1992; Okada et al., 2005), and the activity of both *HMT/HLT* and *ECT/EFT-LCT/LFT* was not associated with chloroplasts (Suzuki et al., 1996). This suggests that while the most important steps in the QA biosynthetic pathway take place in aerial tissues, it is possible that the entire pathway is not limited to such tissues. Once synthesized, QAs are then translocated to the reproductive organs via the phloem (Wink and Witte, 1984; Lee et al., 2007). The loading of QAs onto the phloem may be selective as lupin leaves have more diverse QA profiles than both grain and phloem exudates (Wink et al., 1995; Lee et al., 2007). No studies have yet investigated QA biosynthesis within seeds themselves. It has been estimated, based on measures of translocation of QAs and total QAs in reproductive tissues, that of the QAs that accumulate in seeds, half are synthesized within the seed and half are translocated (Lee et al., 2007).

The identification of sites of QA biosynthesis and transport processes is important for targeting the accumulation of QAs in grain. If QA biosynthesis within seeds themselves is not appreciable, this offers the means to target QA transport processes in order to reduce grain QA levels without compromising QA biosynthetic processes, which negatively affects plant fitness. In lupin, sweet (low QA) cultivars have considerably lower resistance to disease and predation compared to bitter (high QA) wild germplasm, increasing susceptibility to insect attack and transmission of aphid-borne viruses (Berlandier, 1996; Wang et al., 2000; Adhikari et al., 2012). In particular, sweet *L. luteus* varieties, which are valued because of their very high protein content, are susceptible to aphid attack and as such, are unsuccessful in Australia and resistance may be difficult to achieve with grain QA levels below 0.02% (Berlandier and Sweetingham, 2003; Adhikari, 2007). One concept for lupin breeding is to develop a 'bitter/sweet' phenotype—a plant that has sufficiently high QA levels in vegetative tissues to deter insect attack, but contains low QA levels in grain (Wink, 1990, 1994). For this, transporters involved in the translocation of QAs from source tissues to seeds must be identified.

Though several genes that are involved in the transport of nicotine, BIAs and a MIA precursor have been identified, no targets are yet identified which affect the alkaloid levels in source and sink tissues separately. In *Nicotiana*, nicotine is synthesized in roots and is usually transported to leaves *via* the xylem (Dawson, 1942; Baldwin, 1989). Transporters involved in the sequestration of nicotine into vacuoles belong to the multidrug and toxic compound extrusion (MATE) family (*NtMATE1*, *NtMATE2*, *NtJAT1*, *NtJAT2*; Morita et al., 2009; Shoji et al., 2009; Shitan et al., 2014) and a plasma membrane located,

nicotine importer belongs to the purine uptake permease-like (PUP-like) family (*NtNUP1*; Hildreth et al., 2011). These nicotine transporters are mainly expressed in roots (*NtMATE1*, *NtMATE2*, *NtNUP1*), with *NtJAT1* expressed in all tissues and *NtJAT2* expressed almost exclusively in leaves and all are induced by methyl jasmonate (Morita et al., 2009; Shoji et al., 2009; Hildreth et al., 2011; Kato et al., 2014; Shitan et al., 2014). *NtJAT1* may also function as a plasma membrane localized nicotine efflux transporter when produced in root tissue, suggesting that this transporter plays more than one key role in nicotine transport (Morita et al., 2009). Most of these MATE transporters also efficiently transport tropane alkaloids, with *NtJAT1* and *NtJAT2* additionally found to transport berberine, and *NtNUP1* also transports vitamin B6 (Morita et al., 2009; Shoji et al., 2009; Hildreth et al., 2011; Shitan et al., 2014; Kato et al., 2015). Down-regulation of *NtMATE1/MATE2* transcript levels in *Nicotiana* plants using RNA-interference (RNAi) did not affect alkaloid levels in the leaves or the roots, however, did increase sensitivity of the plant to exogenously applied nicotine (Shoji et al., 2009). Down-regulation of *NtNUP1* reduced nicotine accumulation throughout the entire plant, however, root to shoot translocation was unaffected (Hildreth et al., 2011; Kato et al., 2014). Interestingly, *NtNUP1* positively regulates the expression of a key transcription factor in the nicotine biosynthesis pathway, possibly explaining the reduced nicotine content in RNAi lines (Kato et al., 2014). It would be interesting to assess the effect of down-regulating *NtJAT1* and *NtJAT2*, as these function as nicotine transporters in sink tissues (Morita et al., 2009; Shitan et al., 2014) and perhaps nicotine levels in leaf tissues may be reduced, while levels in roots may be increased or unaffected.

One *Nicotiana* species—*N. alata*—synthesizes nicotine in the roots but is unable to translocate it to the xylem for transport to the leaves (Pakdeechanuan et al., 2012). Genetic studies involving hybrids between *N. alata* and the closely related *N. langsdorffii*, which does accumulate nicotine in leaf tissue, indicate that more than one dominant locus is involved in blocking transport of nicotine from the root to the xylem (Pakdeechanuan et al., 2012). The expression of *NtMATE1* and *NtMATE2* is also observed in the root (Pakdeechanuan et al., 2012). The identification of those genes controlling the dominant loci blocking nicotine transport will further our understanding of the long-distance transport of plant alkaloids.

In *C. japonica*, berberine is transported from the lateral roots to the rhizome (Fujiwara et al., 1993). Three berberine transporters have been identified and belong to the ATP-binding cassette (ABC) family (*CjABCB1/CjMDR1*, *CjABCB2*, and *CjABCB3*), with *CjABCB1* and *CjABCB2* localized in the plasma membrane and expressed in the rhizome, possibly playing a role in the uptake of berberine in the rhizome (Shitan et al., 2003, 2013). In transgenic *C. japonica*, where *CjABCB1* was suppressed, berberine accumulation in the root decreased (Shitan et al., 2005).

In *C. roseus*, catharanthine—which is coupled with vindoline to produce vinblastine and vincristine—is transported from the leaf epidermis to the leaf surface, resulting in spatial separation of catharanthine and vindoline (Roepke et al., 2010). An ABC transporter, CrTPT2, which is specifically expressed in the leaf

epidermis, functions as a catharanthine exporter (Yu and De Luca, 2013). Virus induced gene silencing (VIGS) of the *CrTPT2* in *C. roseus* resulted in reduced catharanthine levels on the leaf surface and caused an increased in catharanthine-vindoline dimers within leaves, demonstrating that altered transport of MIA intermediates may alter biosynthesis of MIAs (Yu and De Luca, 2013).

It is evident that altered expression of alkaloid transporters is able to alter the accumulation of alkaloids, whether that be through changes in transport processes and/or regulation of alkaloid biosynthesis itself. Candidate transporters for altering alkaloid accumulation processes would, however, need a high degree of specificity in recognizing the target alkaloids in order to not alter other transport processes in the plant, which may have undesirable consequences. In the case of QA transport in lupin, the transporters involved would include plasma membrane located exporters in cells of aerial tissue, membrane-localized transporters for entry onto the phloem, plasma membrane importers in cells of reproductive tissue, and vacuolar membrane importers in cells of both aerial and reproductive tissue, as alkaloids are often sequestered within vacuoles to avoid toxic effects within tissues (Yazaki et al., 2008). The transporters of most interest for lupin breeding would be those involved in the import of QAs into the grain from the phloem, as QA levels in aerial tissue and phloem sap would ideally be high to deter feeding of chewing and sap-sucking insects.

GENES CONTROLLING QA CONTENT IN LUPINS

Lupin 'Low Alkaloid' Domestication Genes

In addition to QA biosynthetic genes, major loci controlling QA content are known. All modern lupin cultivars display a significantly lower QA phenotype compared with wild varieties due to 'low-alkaloid' domestication genes, specific for each lupin species with most arising from natural mutation. Low-alkaloid mutants of NLL, *L. luteus*, *L. albus*, and *L. mutabilis* were first identified in Germany in the late 1920s to early 1930s from wild germplasm (von Sengbusch, 1942) and give insights into the regulation of QA biosynthesis.

Natural low alkaloid mutations in NLL (*iucundus*, *esculentus*, and *depressus*) and *L. luteus* (*amoenus*, *dulcis*, and *liber*) are recessive, segregating independently of one another and follow a simple Mendelian inheritance pattern of clear 1:3 segregation (von Sengbusch, 1942; Gustafsson and Gadd, 1965). A fourth NLL locus *tantalus* was later identified by x-ray induced mutation (Zachow, 1967). The locus *iucundus* appears to have been exclusively used for NLL breeding and *dulcis* for *L. luteus* breeding (Lamberts, 1955; Gustafsson and Gadd, 1965; Gladstones, 1970). Of various, presumed natural, recessive low-alkaloid mutations in *L. albus*, identified by several authors, *pauper*, *mitis*, *reductus*, *exiguus*, and *nutricius* are located at different loci (Harrison and Williams, 1982; Kurlovich, 2002). The *pauper* locus is the most effective mutation in reducing QA

levels and is now almost exclusively used in breeding programs, though *nutricius* and *exiguus* have been used in certain cultivars (Gladstones, 1970; Harrison and Williams, 1982). Low alkaloid material of *L. mutabilis* identified in the 1930s was lost (von Sengbusch, 1942), and it was not until other natural low alkaloid mutants were reselected over several generations, that the first sweet variety with grain content less than 0.05% was developed in the early 1980s (von Baer and Gross, 1977, 1983; Clements et al., 2008). The cultivar Inti was then produced which has a QA content less than 0.02% (Gross et al., 1988). Crosses between Inti and bitter *L. mutabilis* revealed inheritance of the low alkaloid trait is recessive, though F2 segregation is slightly higher than 1:4, indicating that the low alkaloid phenotype in Inti is controlled by a major, as well as additional minor alleles (Clements et al., 2008). Of the low-alkaloid mutations identified in lupin, none eliminates QAs completely (Gustafsson and Gadd, 1965; Gladstones, 1970; Harrison and Williams, 1982).

For low-alkaloid loci *dulcis* and presumably *pauper*, the limiting step of the QA pathway may be the reaction from cadaverine to the major structural QAs, as lysine and cadaverine levels do not differ between sweet and bitter plants in *L. luteus* and *L. albus*, nor do enzymatic activities for QA acyltransferases (Saito et al., 1993). Sweet NLL harboring the *iucundus* locus appears to have lower levels of lysine than bitter wild NLL, suggesting a different function for this gene (Bunsupa et al., 2012a). While these species of lupin cannot be crossed, the identification of low alkaloid genes will assist in further elucidating the QA pathway and will allow homologous genes between species to be identified and targeted in breeding programs. Markers linked to low alkaloid loci *iucundus* and *pauper* have been developed to assist tracking these recessive loci (Lin et al., 2009; Li et al., 2011). Recently, dense mapping resources, an updated genetic map for NLL cv. Tanjil and genome annotation have further narrowed the candidate gene region of *iucundus* on NLL-07 (Hane et al., 2016).

Regulators of Alkaloid Biosynthesis

Jasmonic acid (JA) is a plant hormone regulating defense responses against environmental stresses and attack by pathogens and insects (Farmer et al., 2003) and is a well-known activator of alkaloid biosynthesis in *Nicotiana*, *C. roseus* and *C. japonica*; the expression of biosynthetic genes and alkaloid levels in nicotine, MIA and BIA biosynthesis, as well as previously mentioned *Nicotiana* and *C. roseus* alkaloid transporters, respond positively to JA treatment (Aerts et al., 1994; Pauli and Kutchan, 1998; Shoji et al., 2008, 2009; Morita et al., 2009; Yu and De Luca, 2013; Shitan et al., 2014; Gurkok et al., 2015; Kato et al., 2015). Many transcription factors regulating alkaloid biosynthesis, including basic Helix-Loop-Helix (bHLH), AP2/ERF and WRKY transcription factors identified in *Nicotiana*, *C. roseus*, and *C. japonica*, also show JA responsiveness (Menke et al., 1999; Chatel et al., 2003; Kato et al., 2007; Shoji et al., 2010; Todd et al., 2010; Suttipanta et al., 2011; Yamada et al., 2011; Van Moerkercke et al., 2015). These bHLH and AP2/ERF transcription factors regulate alkaloid biosynthesis by recognizing GCC-motif and G-box elements in the promoters of alkaloid biosynthetic genes

in *Nicotiana* and *C. roseus* (Chatel et al., 2003; Shoji et al., 2010; De Boer et al., 2011), while a WRKY transcription factor binds to W-box elements in a *C. roseus* alkaloid biosynthetic gene promoter (Suttipanta et al., 2011). A *C. roseus* bHLH transcription factor can also bind to a G-box-like element in an AP2/ERF promoter which in turn promotes MIA biosynthesis (Zhang et al., 2011). As QA levels in lupin vegetative material is known to increase after wounding (Wink, 1983; Chludil et al., 2013), QA biosynthesis may be regulated by the JA pathway in similar ways to other alkaloids and it may be possible to identify similar candidate transcription factors regulating QA biosynthesis. More recently, microRNAs (miRNAs)—endogenous, small, non-coding RNAs which regulate gene expression by causing target mRNA degradation or translational repression (Naqvi et al., 2012)—have been identified which may regulate nicotine and BIA biosynthetic genes (Boke et al., 2015; Li et al., 2015). Nicotine biosynthesis can also be controlled by non-coding target mimicry (eTM)-mediated inhibition of its corresponding nicotine biosynthetic gene-targeting miRNA (Li et al., 2015). It will therefore be interesting to analyze the role of miRNAs in regulating QA biosynthesis in lupin. Putative miRNAs have been identified from *L. albus* phloem exudate (Rodriguez-Medina et al., 2011), and as additional lupin miRNA data sets become available, miRNAs regulating QA biosynthesis may be identified in order to better understand how the pathway is regulated.

ENVIRONMENTAL FACTORS AFFECTING QA PRODUCTION

There is a significant environmental impact on grain QA content in lupin, due to either regulation of QA biosynthesis or transport from source tissues to the seed, though this impact appears to be highly unpredictable, with QA levels poorly explained by environmental properties such as location and seasonal climate (Cowling and Tarr, 2004). Grain QA levels can often exceed industry limits, usually by a couple of fold, though concentrations up to 2120 mg/kg have been found in sweet NLL, exceeding the limit by more than 10 times (Cowling and Tarr, 2004; Reinhard et al., 2006). There is, therefore, a great need to better understand how QA biosynthesis and transport is affected by environmental factors.

Light regulates QA biosynthesis by affecting the conditions within the chloroplast, with L/ODC activated by reduced thioredoxin and a light-mediated shift in pH of the chloroplast stroma from pH 7 to 8 during the day (Wink and Hartmann, 1981). As such, QA biosynthesis displays a diurnal rhythm whereby leaf QA concentrations are higher during the day and lower during the night (Wink and Witte, 1984). As light conditions cannot be controlled in the field, this factor is likely of less concern to breeders and farmers. Drought conditions are thought to increase QA content in lupin and drought stress can increase alkaloid levels in *Nicotiana*, *P. somniferum*, and *C. roseus* (Waller and Nowacki, 1978; Szabó et al., 2003; Jaleel et al., 2007). The effect of drought on grain QA content in lupin is not clear, with the plant growth stage at which drought is imposed

seeming to play a role in whether QA content increases or decreases, albeit marginally (Christiansen et al., 1997), however, the amount of rainfall is not strongly associated with seed QA content (Cowling and Tarr, 2004). Ambient temperature seems to have an important effect on QA content, with a small increase in mean temperature (3°C) having a marked increase in grain QA content in European NLL varieties (Jansen et al., 2009, 2015). Soil characteristics, such as soil pH and the type and amount of fertilizer used, also affect grain QA levels. Higher soil pH (6.7 and 7.1) results in lower QAs than lower soil pH (5.3 and 5.8; Jansen et al., 2012). Potassium deficiency increases QAs, while phosphorus deficiency reduces them, with a significant interaction between potassium and phosphorous on QA content (Gremigni et al., 2001, 2003). The growing system also has a small effect on grain QA content, with organic conditions resulting in lower QA content than conventional conditions (Jansen et al., 2015). The amplitude of the response of grain QA content to environmental factors is also dependent on genotype, with some NLL cultivars more variable in QA content than others (Gremigni et al., 2000; Cowling and Tarr, 2004; Jansen et al., 2015).

While a few studies have investigated the role of abiotic stresses on QA biosynthesis, there are currently no reports of the impact of biotic stresses on grain QA content. As QAs play a role in the protection of the plant from predators, it is thought that QA accumulation may increase as part of a defense response when the plant comes under attack. Mechanical wounding of lupin leaves and plants, which may mimic herbivore action, has increased QA accumulation in vegetative material (Wink, 1983; Chludil et al., 2013). Leaf damage also leads to an increase in nicotine biosynthesis in *Nicotiana* (Baldwin, 1989; Cane et al., 2005). In a field situation, however, the large-scale wounding of lupin crops is not likely. Of greater concern to lupin growers are insect pests such as aphids, which can cause significant yield losses (Berlandier and Sweetingham, 2003). While it is known that QAs are a feeding deterrent to aphids (Berlandier, 1996; Ridsdill-Smith et al., 2004; Adhikari et al., 2012; Philippi et al., 2015), how QA production, and QA content in lupin grain respond to aphid attack has not yet been investigated. Additionally, the attack of lupin plants by fungal pathogens may impact QA production as alkaloid biosynthetic genes in *C. roseus* and *P. somniferum* are induced by treatment with fungal elicitors (Pasquali et al., 1992; Facchini et al., 1996a). It is likely that many different factors impact on QA biosynthesis in a field situation, and those most important need to be identified in order to be able to grow a valuable lupin crop.

QA QUANTIFICATION METHODS

High or low grain QA phenotypes in lupin were first identified by a method similar to the still-employed Dragendorff test; the Dragendorff reagent reacts with high QA phenotypes (>0.5%), described as bitter, and low alkaloid phenotypes with no reaction are described as sweet (Harrison and Williams, 1982; Harborne, 1984). More accurate and nowadays the more common method of QA quantification is performed with gas-chromatography (sometimes gas-liquid chromatography)

combined with a detector, usually a mass spectrometer. Most studies report on QA content in lupin grain and food products, though some studies also report on leaves and less commonly flowers, stems, roots and phloem sap (Gremigni et al., 2003; Cowling and Tarr, 2004; Reinhard et al., 2006; Lee et al., 2007; Resta et al., 2008; Hernández et al., 2011; Adhikari et al., 2012; Kamel et al., 2015). QA extraction is performed by leaching compounds from samples using an aqueous acidic solution, and then adjusting the solution to a basic pH for QA extraction with an organic solvent, usually dichloromethane (Wink et al., 1995; Ganzera et al., 2010). Wink et al. (1995) provide the most comprehensive spectral dataset for QAs, reporting Kovat's indices and mass spectral data for 100 alkaloids found in different species of lupin. Many subsequent studies make use of this mass spectral data to identify QAs (Erdemoglu et al., 2007; Resta et al., 2008) as obtaining pure chemical standards of QAs is difficult as they are expensive and not readily available commercially. As a consequence, quantification has been based on relative concentrations of QAs (Erdemoglu et al., 2007; Chludil et al., 2009) or standard curves of major alkaloids (e.g., lupanine, gramine, or sparteine) or internal standards (e.g., caffeine, matrine) which are then applied to estimate concentrations of various other QAs (Muzquiz et al., 1994; Resta et al., 2008). Isolation of reference QAs from lupin tissue is also possible (Brooke et al., 1996; Wang et al., 2000; Reinhard et al., 2006) and has been used to quantify major QAs in NLL grain (Priddis, 1983; Harris and Wilson, 1988). Limit of detection (LOD) and limit of quantification (LOQ) values for the identification and quantification of QAs are not often reported, but for those that are, LODs range from 1 to 30.5 $\mu\text{g/mL}$ and LOQs range from 3 to 87 $\mu\text{g/mL}$ (Table 2) (Boschin et al., 2008; Resta et al., 2008; Ganzera et al., 2010). Despite many reports of quantification of QAs, many studies rely on relative QA quantifications or quantification for certain major QAs in mainly seed material of cultivated lupin species. In addition to the few major QAs, the presence of many minor QAs has been established in lupin leaf and grain (Wink et al., 1995) and levels of these may need further evaluation. Particularly in the case of cultivated NLL, which has a narrow genetic base and for which wild germplasm is often used as a source of genetic variation in pre-breeding material (Cowling et al., 2009; Berger et al., 2012, 2013), additional levels of minor QAs which are not being monitored may be inadvertently introduced by such breeding practices. There is, therefore, a need for an improved and more thorough methodology for the detection and quantification of QAs in lupin grain as well as other tissue types, for monitoring grain QA levels for food safety and to further facilitate the understanding of QA biosynthesis and accumulation.

FUTURE PROSPECTS

There is now a realistic opportunity for further elucidating the QA biosynthetic pathway in lupin grain crops and tackling the problem of QA accumulation for an emerging human health food. Recently developed genetic and genomic resources for NLL will greatly facilitate the identification of genes involved

TABLE 2 | Limit of detection (LOD) and limit of quantification (LOQ) values for measurement of quinolizidine alkaloids by gas-chromatography mass-spectrometry (GC-MS) or capillary-electrophoresis mass-spectrometry (CE-MS) in lupin grain and lupin-based foods.

Compound	LOD ($\mu\text{g/mL}$)	LOQ ($\mu\text{g/mL}$)	Reference
Lupanine	2	3	Boschin et al., 2008; Resta et al., 2008
	7.9	18.2	Ganzera et al., 2010
Sparteine	1	3	Boschin et al., 2008; Resta et al., 2008
	30.5	87	Ganzera et al., 2010
13 α -hydroxylupanine	6.5	12.5	Ganzera et al., 2010
Angustifoline	16.1	61.4	Ganzera et al., 2010

in this pathway, including the generation of a comprehensive NLL genome sequence (Hane et al., 2016), transcriptomic data sets for various NLL tissue types as well as three other cultivated lupin species (Parra-González et al., 2012; O'Rourke et al., 2013; Secco et al., 2014; Wang et al., 2014; Foley et al., 2015; Kamphuis et al., 2015), and dense genetic maps for NLL and *L. albus* (Croxford et al., 2008; Nelson et al., 2010; Yang et al., 2013; Kroc et al., 2014; Kamphuis et al., 2015). The remarkable progress in the elucidation of alkaloid biosynthetic pathways in model species in recent years, using a combination of genetic maps, genomic and transcriptomic resources, technical advances in enzymology, next generation sequencing, metabolite profiling and methodology for validating candidate genes with roles in alkaloid biosynthesis (i.e., RNAi or VIGS; Hagel and Facchini, 2013; Gurkok et al., 2015; Wang and Bennetzen, 2015; Kilgore and Kutchan, 2016; Pan et al., 2016), serves as a strong basis for understanding QA biosynthesis. Genetic and genomic resources can now be utilized in lupin to identify transcriptome-based candidate genes involved in QA biosynthesis and transport by comparative analysis between high and low QA varieties and plant tissue types or transcriptomic profiling analyzing QA-induced plants. The function of candidate genes may be studied by genetic transformation of lupin, the primary method being *Agrobacterium tumefaciens*-mediated transformation of wounded seedling shoot apical meristems to generate transgenic shoots (Pigeaire et al., 1997). While transformation efficiencies are low due to low survival and chimeric nature of T_0 plants, this method has been successful in generating transgenic NLL (Molvig et al., 1997; Pigeaire et al., 1997; Wijayanto et al., 2009; Tabe et al., 2010; Atkins et al., 2011; Barker et al., 2016), *L. luteus* (Li et al., 2000; Pniewski et al., 2006), and *L. mutabilis* (Babaoglu et al., 2000; Polowick et al., 2014) to confer various traits, with recent modifications improving this transformation method for NLL (Nguyen et al., 2016a,b). For *L. albus*, however, *A. tumefaciens*-mediated transformation has been unsuccessful and as such hairy root transformation using *A. rhizogenes* (Uhde-Stone et al., 2005; Sbabou et al., 2010; Cheng et al., 2011) and VIGS using the *Peanut stunt virus* vector (Yamagishi et al., 2015) have been used to study gene function in this species. Metabolite profiling is an additional resource that could be enhanced in lupins to provide a valuable understanding of how the QA pathway interacts with other metabolic pathways in the plant,

especially under abiotic and biotic stresses. This would be useful in understanding the effect of environmental and genotypic interactions on QA biosynthesis. The metabolite profiling of genetically diverse wild NLL accessions that will be used to introgress novel traits into pre-breeding lines would also be of interest as the level of genetic variation in QA content and composition in wild NLL is unclear and accessions which may introduce QAs which are currently not monitored for, need to be identified.

A better understanding of the genes involved in QA production and transport will allow for the management of QA grain content in various ways. For NLL, the first approach would be to focus on introgressing recessive alleles other than the *iucundus* locus (e.g., *esculentus*, *depressus*, and *tantalus*) into new varieties with the hypothesis that stacking these could reduce the QA content further and well-below the 0.02% threshold for use as a product for human consumption. A second approach would be to use various reverse-genetics approaches to identify genes involved in the biosynthesis, regulation or transport of QAs in order to reduce QA biosynthesis or transport to the grain. As most reduced QA content in lupins are the result of spontaneous mutations, and low QA loci are simply inherited and thus major genes controlling the pathway, it may be possible to find complete knock out mutants in QA biosynthetic genes as none of the current low QA mutations remove the QA phenotype completely. The targeting of QA transporter genes may also allow the QA content in foliar tissue to remain at high levels, but reduce or nullify the transport of QAs to the seed thereby still providing strong protection of the foliar tissue to insect predation, yet producing grain suitable for human consumption. A third approach would be the use of CRISPR/Cas9 to edit genes involved in QA regulation, synthesis or transport, thereby reducing grain

QA content. The use of this technology will depend on whether its products are classified as genetically modified in the regulatory systems of different countries. In Australia, where the majority of the world's lupin grain is produced, current legislation would class this as genetically modified, and CRISPR/Cas9 is therefore a less desirable approach for the improvement of lupin crops.

In stark contrast to most other crop species, lupins are only very recently domesticated and modern varieties have a narrow genetic base. The excellent genetic and genomic resources available for lupin now offer significant opportunities to ensure grain QA levels remain below the industry limit to improve the quality of this high protein grain legume.

AUTHOR CONTRIBUTIONS

KF wrote the manuscript with input from RF, KHMS, KBS, and LK.

FUNDING

This project is supported by a University Postgraduate Award of the University of Western Australia and a Grains Research and Development Corporation (GRDC) Scholarship (GRS10935) awarded to KF.

ACKNOWLEDGMENT

We thank Kathleen de Boer and Lingling Gao for helpful comments on the manuscript.

REFERENCES

- Adhikari, K. N. (2007). *Yellow Lupin Improvement: A Final Report Prepared for the Grains Research and Development Corporation: GRDC Project Number DAW00104*. South Perth, WA: Department of Agriculture and Food.
- Adhikari, K. N., Edwards, O. R., Wang, S., Ridsdill-Smith, T. J., and Buirchell, B. (2012). The role of alkaloids in conferring aphid resistance in yellow lupin (*Lupinus luteus* L.). *Crop Pasture Sci.* 63, 444–451. doi: 10.1071/CP12189
- Aerts, R. J., Gisi, D., De Carolis, E., De Luca, V., and Baumann, T. W. (1994). Methyl jasmonate vapor increases the developmentally controlled synthesis of alkaloids in *Catharanthus* and *Cinchona* seedlings. *Plant J.* 5, 635–643. doi: 10.1111/j.1365-3113.1994.00635.x
- Allen, J. G. (1998). "Toxins and lupinosis," in *Lupins as Crop Plants: Biology, Production and Utilization*, eds J. S. Glasstones, C. A. Atkins, and J. Hamblin (Cambridge: CAB International), 411–428.
- Atkins, C., Emery, R., and Smith, P. (2011). Consequences of transforming narrow leaved lupin (*Lupinus angustifolius* [L.]) with an ipt gene under control of a flower-specific promoter. *Transgenic Res.* 20, 1321–1332. doi: 10.1007/s11248-011-9497-7
- Babaoglu, M., McCabe, M. S., Power, J. B., and Davey, M. R. (2000). Agrobacterium-mediated transformation of *Lupinus mutabilis* L. using shoot apical explants. *Acta Physiol. Plant.* 22, 111–119. doi: 10.1007/s11738-000-0064-8
- Baldwin, I. T. (1989). Mechanism of damage-induced alkaloid production in wild tobacco. *J. Chem. Ecol.* 15, 1661–1680. doi: 10.1007/BF01012392
- Barker, S. J., Si, P., Hodgson, L., Ferguson-Hunt, M., Khentry, Y., Krishnamurthy, P., et al. (2016). Regeneration selection improves transformation efficiency in narrow-leaf lupin. *Plant Cell Tissue Organ Cult.* 126, 219–228. doi: 10.1007/s11240-016-0992-7
- Beaudoin, G. A., and Facchini, P. J. (2014). Benzyloquinoline alkaloid biosynthesis in opium poppy. *Planta* 240, 19–32. doi: 10.1007/s00425-014-2056-8
- Berger, J. D., Buirchell, B. J., Luckett, D. J., and Nelson, M. N. (2012). Domestication bottlenecks limit genetic diversity and constrain adaptation in narrow-leaved lupin (*Lupinus angustifolius* L.). *Theor. Appl. Genet.* 124, 637–652. doi: 10.1007/s00122-011-1736-z
- Berger, J. D., Clements, J. C., Nelson, M. N., Kamphuis, L. G., Singh, K. B., and Buirchell, B. (2013). The essential role of genetic resources in narrow-leaved lupin improvement. *Crop Pasture Sci.* 64, 361–373. doi: 10.1071/CP13092
- Berlandier, F. A. (1996). Alkaloid level in narrow-leaved lupin, *Lupinus angustifolius*, influences green peach aphid reproductive performance. *Entomol. Exp. Appl.* 79, 19–24. doi: 10.1111/j.1570-7458.1996.tb00804.x
- Berlandier, F. A., and Sweetingham, M. W. (2003). Aphid feeding damage causes large losses in susceptible lupin cultivars. *Anim. Prod. Sci.* 43, 1357–1362. doi: 10.1071/EA02186
- Bird, D. A., Franceschi, V. R., and Facchini, P. J. (2003). A tale of three cell types: alkaloid biosynthesis is localized to sieve elements in opium poppy. *Plant Cell* 15, 2626–2635. doi: 10.1105/tpc.015396
- Boke, H., Ozhuner, E., Turktas, M., Parmaksiz, I., Ozcan, S., and Unver, T. (2015). Regulation of the alkaloid biosynthesis by miRNA in opium poppy. *Plant Biotechnol. J.* 13, 409–420. doi: 10.1111/pbi.12346
- Boschin, G., Annicchiarico, P., Resta, D., D'Agostina, A., and Arnoldi, A. (2008). Quinolizidine alkaloids in seeds of lupin genotypes of different origins. *J. Agric. Food Chem.* 56, 3657–3663. doi: 10.1021/jf7037218

- Boschin, G., and Resta, D. (2013). "Alkaloids derived from lysine: quinolizidine (a focus on lupin alkaloids)," in *Natural Products*, eds K. G. Ramawat and J.-M. Merillon (Berlin: Springer), 381–403. doi: 10.1007/978-3-642-22144-6
- Brooke, P., Harris, D. J., and Longmore, R. B. (1996). Isolation of minor lupin alkaloids. 1. A simple procedure for the isolation of angustifoline from *Lupinus angustifolius* (cv. Fest) seeds, with application to other lupin alkaloids. *J. Agric. Food Chem.* 44, 2129–2133. doi: 10.1021/jf950020y
- Bunsupa, S., Hanada, K., Maruyama, A., Aoyagi, K., Komatsu, K., Ueno, H., et al. (2016). Molecular evolution and functional characterization of a bifunctional decarboxylase involved in Lycopodium alkaloid biosynthesis. *Plant Physiol.* 171, 2432–2444. doi: 10.1104/pp.16.00639
- Bunsupa, S., Katayama, K., Ikeura, E., Oikawa, A., Toyooka, K., Saito, K., et al. (2012a). Lysine decarboxylase catalyzes the first step of quinolizidine alkaloid biosynthesis and coevolved with alkaloid production in Leguminosae. *Plant Cell* 24, 1202–1216. doi: 10.1105/tpc.112.095885
- Bunsupa, S., Okada, T., Saito, K., and Yamazaki, M. (2011). An acyltransferase-like gene obtained by differential gene expression profiles of quinolizidine alkaloid-producing and nonproducing cultivars of *Lupinus angustifolius*. *Plant Biotechnol.* 28, 89–94. doi: 10.5511/plantbiotechnology.10.1109b
- Bunsupa, S., Yamazaki, M., and Saito, K. (2012b). Quinolizidine alkaloid biosynthesis: recent advances and future prospects. *Front. Plant Sci.* 3:239. doi: 10.3389/fpls.2012.00239
- Cane, K. A., Mayer, M., Lidgett, A. J., Michael, A. J., and Hamill, J. D. (2005). Molecular analysis of alkaloid metabolism in AAB v. aabb genotype *Nicotiana tabacum* in response to wounding of aerial tissues and methyl jasmonate treatment of cultured roots. *Funct. Plant Biol.* 32, 305–320. doi: 10.1071/FP04008
- Chatel, G., Montiel, G., Pré, M., Memelink, J., Thiersault, M., Saint-Pierre, B., et al. (2003). CrMYC1, a *Catharanthus roseus* elicitor- and jasmonate-responsive bHLH transcription factor that binds the G-box element of the strictosidine synthase gene promoter. *J. Exp. Bot.* 54, 2587–2588. doi: 10.1093/jxb/erg275
- Cheng, L., Bucciarelli, B., Liu, J., Zinn, K., Miller, S., Patton-Vogt, J., et al. (2011). White lupin cluster root acclimation to phosphorus deficiency and root hair development involve unique glycerophosphodiester phosphodiesterases. *Plant Physiol.* 156, 1131–1148. doi: 10.1104/pp.111.173724
- Chludil, H. D., Leicach, S. R., Corbino, G. B., Barriga, L. G., and Vilariño, M. D. P. (2013). Genistin and quinolizidine alkaloid induction in *L. angustifolius* aerial parts in response to mechanical damage. *J. Plant Interact.* 8, 117–124. doi: 10.1080/17429145.2012.672660
- Chludil, H. D., Vilariño, M. D. P., Franco, M. A. L., and Leicach, S. R. (2009). Changes in *Lupinus albus* and *Lupinus angustifolius* alkaloid profiles in response to mechanical damage. *J. Agric. Food Chem.* 57, 6107–6113. doi: 10.1021/jf901232c
- Christiansen, J. L., Jørgsgård, B., Buskov, S., and Olsen, C. E. (1997). Effect of drought stress on content and composition of seed alkaloids in narrow-leaved lupin, *Lupinus angustifolius* L. *Eur. J. Agron.* 7, 307–314. doi: 10.1016/S1161-0301(97)00017-8
- Clements, J. C., Sweetingham, M. S., Smith, L., Francis, G., Thomas, G., and Sipsas, S. (2008). "Crop improvement in *Lupinus mutabilis* for Australian agriculture-Progress and prospects," in *Proceedings of the 12th International Lupin Conference* (Geraldton: International Lupin Association), 14–18.
- Cowling, W., Buirchell, B., and Falk, D. (2009). A model for incorporating novel alleles from the primary gene pool into elite crop breeding programs while reselecting major genes for domestication or adaptation. *Crop Pasture Sci.* 60, 1009–1015. doi: 10.1071/CP08223
- Cowling, W. A., Huyghe, C., and Swiecicki, W. (1998). "Lupin breeding," in *Lupinus as Crop Plants: Biology, Production and Utilization*, eds J. S. Glasstones, C. A. Atkins, and J. Hamblin (Cambridge: CAB International), 93–120.
- Cowling, W. A., and Tarr, A. (2004). Effect of genotype and environment on seed quality in sweet narrow-leaved lupin (*Lupinus angustifolius* L.). *Aust. J. Agric. Res.* 55, 745–751. doi: 10.1071/AR03223
- Croxford, A. E., Rogers, T., Caligari, P. D., and Wilkinson, M. J. (2008). High-resolution melt analysis to identify and map sequence-tagged site anchor points onto linkage maps: a white lupin (*Lupinus albus*) map as an exemplar. *New Phytologist* 180, 594–607. doi: 10.1111/j.1469-8137.2008.02588.x
- Daverio, M., Cavicchiolo, M. E., Grotto, P., Lonati, D., Cananzi, M., and Da Dalt, L. (2013). Bitter lupine beans ingestion in a child: a disregarded cause of acute anticholinergic toxicity. *Eur. J. Pediatr.* 173, 1549–1551. doi: 10.1007/s00431-013-2088-2
- Dawson, R. F. (1942). Accumulation of nicotine in reciprocal grafts of tomato and tobacco. *Am. J. Bot.* 29, 66–71. doi: 10.2307/2436544
- De Boer, K., Tillemann, S., Pauwels, L., Vanden Bossche, R., De Sutter, V., Vanderhaeghen, R., et al. (2011). APETALA2/ETHYLENE RESPONSE FACTOR and basic helix–loop–helix tobacco transcription factors cooperatively mediate jasmonate-elicited nicotine biosynthesis. *Plant J.* 66, 1053–1065. doi: 10.1111/j.1365-3113.2011.04566.x
- de Cortes Sánchez, M., Altares, P., Pedrosa, M. M., Burbano, C., Cuadrado, C., Goyoaga, C., et al. (2005). Alkaloid variation during germination in different lupin species. *Food Chem.* 90, 347–355. doi: 10.1016/j.foodchem.2004.04.008
- Dewey, R. E., and Xie, J. (2013). Molecular genetics of alkaloid biosynthesis in *Nicotiana tabacum*. *Phytochemistry* 94, 10–27. doi: 10.1016/j.phytochem.2013.06.002
- Dewick, P. M. (2002). "Alkaloids," in *Medicinal Natural Products: a Biosynthetic Approach*, second Edn, ed. P. M. Dewick (Chichester: John Wiley and Sons, Ltd), 291–403.
- Duranti, M., Consonni, A., Magni, C., Sessa, F., and Scarafoni, A. (2008). The major proteins of lupin seed: characterisation and molecular properties for use as functional and nutraceutical ingredients. *Trends Food Sci. Technol.* 19, 624–633. doi: 10.1016/j.tifs.2008.07.002
- Erdemoglu, N., Ozkan, S., and Tosun, F. (2007). Alkaloid profile and antimicrobial activity of *Lupinus angustifolius* L. alkaloid extract. *Phytochem. Rev.* 6, 197–201. doi: 10.1007/s11101-006-9055-8
- Facchini, P. J., Johnson, A. G., Poupart, J., and De Luca, V. (1996a). Uncoupled defense gene expression and antimicrobial alkaloid accumulation in elicited opium poppy cell cultures. *Plant Physiol.* 111, 687–697. doi: 10.1104/pp.111.3.687
- Facchini, P. J., Penzes, C., Johnson, A. G., and Bull, D. (1996b). Molecular characterization of berberine bridge enzyme genes from opium poppy. *Plant Physiol.* 112, 1669–1677. doi: 10.1104/pp.112.4.1669
- Farmer, E. E., Alméras, E., and Krishnamurthy, V. (2003). Jasmonates and related oxylipins in plant responses to pathogenesis and herbivory. *Curr. Opin. Plant Biol.* 6, 372–378. doi: 10.1016/S1369-5266(03)00045-1
- Foley, R. C., Jimenez-Lopez, J. C., Kamphuis, L. G., Hane, J. K., Melser, S., and Singh, K. B. (2015). Analysis of conglutin seed storage proteins across lupin species using transcriptomic, protein and comparative genomic approaches. *BMC Plant Biol.* 15:106. doi: 10.1186/s12870-015-0485-6
- Fraser, A. M., and Robins, D. J. (1984). Incorporation of chiral [1-2 H] cadaverines into the quinolizidine alkaloids sparteine, lupanine, and angustifoline. *J. Chem. Soc. Chem. Commun.* 22, 1477–1479. doi: 10.1039/C39840001477
- Fujiwara, H., Takeshita, N., Terano, Y., Fitch, J. H., Tsujita, T., Katagiri, Y., et al. (1993). Expression of (S)-scoulerine 9-O-methyltransferase in *Coptis japonica* plants. *Phytochemistry* 34, 949–954. doi: 10.1016/S0031-9422(00)90692-7
- Ganzera, M., Krüger, A., and Wink, M. (2010). Determination of quinolizidine alkaloids in different *Lupinus* species by NACE using UV and MS detection. *J. Pharm. Biomed. Anal.* 53, 1231–1235. doi: 10.1016/j.jpba.2010.05.030
- Gladstones, J. S. (1970). Lupins as crop plants. *Field Crop Abstracts* 23, 123–148.
- Golebiewski, W. M., and Spenser, I. D. (1988). Biosynthesis of the lupine alkaloids. II. Sparteine and lupanine. *Can. J. Chem.* 66, 1734–1748. doi: 10.1139/v88-280
- Gremigni, P., Hamblin, J., Harris, D., and Cowling, W. (2003). The interaction of phosphorus and potassium with seed alkaloid concentrations, yield and mineral content in narrow-leaved lupin (*Lupinus angustifolius* L.). *Plant Soil* 253, 413–427. doi: 10.1023/A:1024828131581
- Gremigni, P., Hamblin, J., Harris, D., Santen, E. V., Wink, M., Weissmann, S., et al. (2000). "Genotype × environment interactions and lupin alkaloids," in *Proceedings of the 9th International Lupin Conference* (Geraldton: International Lupin Association), 362–365.
- Gremigni, P., Wong, M., Edwards, N., Harris, D., and Hamblin, J. (2001). Potassium nutrition effects on seed alkaloid concentrations, yield and mineral content of lupins (*Lupinus angustifolius*). *Plant Soil* 234, 131–142. doi: 10.1023/A:1010576702139
- Gross, R., Von Baer, E., Koch, F., Marquard, R., Trugo, L., and Wink, M. (1988). Chemical composition of a new variety of the Andean lupin (*Lupinus mutabilis* cv. Inti) with low-alkaloid content. *J. Food Compos. Anal.* 1, 353–361. doi: 10.1016/0889-1575(88)90035-X

- Gurkok, T., Turktas, M., Parmaksiz, I., and Unver, T. (2015). Transcriptome profiling of alkaloid biosynthesis in elicitor induced opium poppy. *Plant Mol. Biol. Rep.* 33, 673–688. doi: 10.1007/s11105-014-0772-7
- Gustafsson, A., and Gadd, I. (1965). Mutations and crop improvement. II. The genus *Lupinus* (Leguminosae). *Hereditas* 53, 15–39. doi: 10.1111/j.1601-5223.1965.tb01977.x
- Hagel, J., and Facchini, P. (2013). Benzylisoquinoline alkaloid metabolism: a century of discovery and a brave new world. *Plant Cell Physiol.* 54, 647–672. doi: 10.1093/pcp/pct020
- Han, R., Takahashi, H., Nakamura, M., Bunsupa, S., Yoshimoto, N., Yamamoto, H., et al. (2015). Transcriptome analysis of nine tissues to discover genes involved in the biosynthesis of active ingredients in *Sophora flavescens*. *Biol. Pharm. Bull.* 38, 876–883. doi: 10.1248/bpb.b14-00834
- Hane, J. K., Ming, Y., Kamphuis, L. G., Nelson, M. N., Garg, G., Atkins, C. A., et al. (2016). A comprehensive draft genome sequence for lupin (*Lupinus angustifolius*), an emerging health food: insights into plant–microbe interactions and legume evolution. *Plant Biotechnol. J.* doi: 10.1111/pbi.12615 [Epub ahead of print].
- Harborne, J. B. (1984). “Methods of plant analysis,” in *Phytochemical Methods: A Guide to Modern Techniques of Plant Analysis*, ed. A. J. Harborne (Dordrecht: Springer), 1–36.
- Harris, D. J., and Wilson, P. E. (1988). “A rapid manual method of lupin alkaloid analysis,” in *Proceedings of the 5th International Lupin Conference* (Geraldton: International Lupin Association), 598–601.
- Harrison, J. E., and Williams, W. (1982). Genetical control of alkaloids in *Lupinus albus*. *Euphytica* 31, 357–364. doi: 10.1007/BF00021651
- Hernández, E. M., Rangel, M. L., Corona, A. E., Angel, J. A., López, J. A. S., Sporer, F., et al. (2011). Quinolizidine alkaloid composition in different organs of *Lupinus aschenbornii*. *Rev. Bras. Farmacog.* 21, 824–828. doi: 10.1590/S0102-695X2011005000149
- Hildreth, S. B., Gehman, E. A., Yang, H., Lu, R.-H., Ritesh, K., Harich, K. C., et al. (2011). Tobacco nicotine uptake permease (NUP1) affects alkaloid metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18179–18184. doi: 10.1073/pnas.1108620108
- Jaleel, C. A., Manivannan, P., Kishorekumar, A., Sankar, B., Gopi, R., Somasundaram, R., et al. (2007). Alterations in osmoregulation, antioxidant enzymes and indole alkaloid levels in *Catharanthus roseus* exposed to water deficit. *Colloids Surf. B: Biointerfaces* 59, 150–157. doi: 10.1016/j.colsurfb.2007.05.001
- Jansen, G., Jürgens, H. U., and Ordon, F. (2009). Effects of temperature on the alkaloid content of seeds of *Lupinus angustifolius* cultivars. *J. Agron. Crop Sci.* 195, 172–177. doi: 10.1111/j.1439-037X.2008.00356.x
- Jansen, G., Jürgens, H.-U., Schliephake, E., and Ordon, F. (2012). Effect of the soil pH on the alkaloid content of *Lupinus angustifolius*. *Int. J. Agron.* 2012, 5. doi: 10.1155/2012/269878
- Jansen, G., Jürgens, H.-U., Schliephake, E., Seddig, S., and Ordon, F. (2015). Effects of growing system and season on the alkaloid content and yield of different sweet *L. angustifolius* genotypes. *J. Appl. Bot. Food Qual.* 88, 1–4. doi: 10.5073/JABFQ.2015.088.001
- Kajikawa, M., Shoji, T., Kato, A., and Hashimoto, T. (2011). Vacuole-localized berberine bridge enzyme-like proteins are required for a late step of nicotine biosynthesis in tobacco. *Plant Physiol.* 155, 2010–2022. doi: 10.1104/pp.110.170878
- Kamel, K. A., Świąćicki, W., Kaczmarek, Z., and Barzyk, P. (2015). Quantitative and qualitative content of alkaloids in seeds of a narrow-leaved lupin (*Lupinus angustifolius* L.) collection. *Genet. Resour. Crop Evol.* 63, 711–719. doi: 10.1007/s10722-015-0278-7
- Kamphuis, L. G., Hane, J. K., Nelson, M. N., Gao, L., Atkins, C. A., and Singh, K. B. (2015). Transcriptome sequencing of different narrow-leaved lupin tissue types provides a comprehensive uni-gene assembly and extensive gene-based molecular markers. *Plant Biotechnol. J.* 13, 14–25. doi: 10.1111/pbi.12229
- Kato, K., Shitan, N., Shoji, T., and Hashimoto, T. (2015). Tobacco NUP1 transports both tobacco alkaloids and vitamin B6. *Phytochemistry* 113, 33–40. doi: 10.1016/j.phytochem.2014.05.011
- Kato, K., Shoji, T., and Hashimoto, T. (2014). Tobacco nicotine uptake permease regulates the expression of a key transcription factor gene in the nicotine biosynthesis pathway. *Plant Physiol.* 166, 2195–2204. doi: 10.1104/pp.114.251645
- Kato, N., Dubouzet, E., Kokabu, Y., Yoshida, S., Taniguchi, Y., Dubouzet, J. G., et al. (2007). Identification of a WRKY protein as a transcriptional regulator of benzylisoquinoline alkaloid biosynthesis in *Coptis japonica*. *Plant Cell Physiol.* 48, 8–18. doi: 10.1093/pcp/pcl041
- Kilgore, M. B., and Kutchan, T. M. (2016). The Amaryllidaceae alkaloids: biosynthesis and methods for enzyme discovery. *Phytochem. Rev.* 15, 317–337. doi: 10.1007/s11101-015-9451-z
- Kroc, M., Koczyk, G., Świąćicki, W., Kilian, A., and Nelson, M. N. (2014). New evidence of ancestral polyploidy in the Genistoid legume *Lupinus angustifolius* L. (narrow-leaved lupin). *Theor. Appl. Genet.* 127, 1237–1249. doi: 10.1007/s00122-014-2294-y
- Kurlovich, B. S. (2002). *Lupins (Geography, Classification, Genetic Resources and Breeding)*. St Petersburg: OY International North Express.
- Lamberts, H. (1955). Broadening the bases for the breeding of yellow sweet lupine. *Euphytica* 4, 97–106. doi: 10.1007/BF00039634
- Lee, M. J., Pate, J. S., Harris, D. J., and Atkins, C. A. (2007). Synthesis, transport and accumulation of quinolizidine alkaloids in *Lupinus albus* L. and *L. angustifolius* L. *J. Exp. Bot.* 58, 935–946. doi: 10.1093/jxb/erl254
- Leistner, E., and Spenser, I. D. (1973). Biosynthesis of the piperidine nucleus. Incorporation of chirally labeled cadaverine-1-3H. *J. Am. Chem. Soc.* 95, 4715–4725. doi: 10.1021/ja00795a041
- Li, F., Wang, W., Zhao, N., Xiao, B., Cao, P., Wu, X., et al. (2015). Regulation of nicotine biosynthesis by an endogenous target mimicry of miRNA in tobacco. *Plant Physiol.* 169, 1062–1071. doi: 10.1104/pp.15.00649
- Li, H., Wylie, S. J., and Jones, M. G. K. (2000). Transgenic yellow lupin (*Lupinus luteus*). *Plant Cell Rep.* 19, 634–637. doi: 10.1007/s002990050785
- Li, X., Yang, H., Buirchell, B., and Yan, G. (2011). Development of a DNA marker tightly linked to low-alkaloid gene iucundus in narrow-leaved lupin (*Lupinus angustifolius* L.) for marker-assisted selection. *Crop Pasture Sci.* 62, 218–224. doi: 10.1071/CP10352
- Lin, R., Renshaw, D., Luckett, D., Clements, J., Yan, G., Adhikari, K., et al. (2009). Development of a sequence-specific PCR marker linked to the gene “pauper” conferring low-alkaloids in white lupin (*Lupinus albus* L.) for marker assisted selection. *Mol. Breed.* 23, 153–161. doi: 10.1007/s11032-008-9222-2
- Mazelis, M., Miflin, B., and Pratt, H. M. (1976). A chloroplast-localized diaminopimelate decarboxylase in higher plants. *FEBS Lett.* 64, 197–200. doi: 10.1016/0014-5793(76)80282-7
- Menke, F. L. H., Champion, A., Kijne, J. W., and Memelink, J. (1999). A novel jasmonate-and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate-and elicitor-inducible AP2-domain transcription factor, ORCA2. *EMBO J.* 18, 4455–4463. doi: 10.1093/emboj/18.16.4455
- Molvig, L., Tabe, L. M., Eggum, B. O., Moore, A. E., Craig, S., Spencer, D., et al. (1997). Enhanced methionine levels and increased nutritive value of seeds of transgenic lupins (*Lupinus angustifolius* L.) expressing a sunflower seed albumin gene. *Proc. Natl. Acad. Sci. U.S.A.* 94, 8393–8398. doi: 10.1073/pnas.94.16.8393
- Morita, M., Shitan, N., Sawada, K., Van Montagu, M. C., Inzé, D., Rischer, H., et al. (2009). Vacuolar transport of nicotine is mediated by a multidrug and toxic compound extrusion (MATE) transporter in *Nicotiana tabacum*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 2447–2452. doi: 10.1073/pnas.0812512106
- Muzquiz, M., Cuadrado, C., Ayet, G., de la Cuadra, C., Burbano, C., and Osagie, A. (1994). Variation of alkaloid components of lupin seeds in 49 genotypes of *Lupinus albus* from different countries and locations. *J. Agric. Food Chem.* 42, 1447–1450. doi: 10.1021/jf00043a011
- Naqvi, A. R., Sarwat, M., Hasan, S., and Roychodhury, N. (2012). Biogenesis, functions and fate of plant microRNAs. *J. Cell. Physiol.* 227, 3163–3168. doi: 10.1002/jcp.24052
- Nelson, M. N., Moolhuijzen, P. M., Boersma, J. G., Chudy, M., Lesniewska, K., Bellgard, M., et al. (2010). Aligning a new reference genetic map of *Lupinus angustifolius* with the genome sequence of the model legume, *Lotus japonicus*. *DNA Res.* 17, 73–83. doi: 10.1093/dnares/dsq001
- Nguyen, A. H., Hodgson, L. M., Erskine, W., and Barker, S. J. (2016a). An approach to overcoming regeneration recalcitrance in genetic transformation of lupins and other legumes. *Plant Cell Tissue Organ Cult.* 127, 623–635. doi: 10.1007/s11240-016-1087-1
- Nguyen, A. H., Wijayanto, T., Erskine, W., and Barker, S. J. (2016b). Using green fluorescent protein sheds light on *Lupinus angustifolius* L. transgenic shoot

- development. *Plant Cell Tissue Organ Cult.* 127, 665–674. doi: 10.1007/s11240-016-1079-1
- Ohmiya, S., Saito, K., and Murakoshi, I. (1995). “Chapter 1 lupine alkaloids,” in *The Alkaloids: Chemistry and Pharmacology*, ed. G. A. Cordell (San Diego, CA: Academic Press), 1–114.
- Okada, T., Hirai, M. Y., Suzuki, H., Yamazaki, M., and Saito, K. (2005). Molecular characterization of a novel quinolizidine alkaloid O-Tigloyltransferase: cDNA cloning, catalytic activity of recombinant protein and expression analysis in *Lupinus* plants. *Plant Cell Physiol.* 46, 233–244. doi: 10.1093/pcp/pci021
- O'Rourke, J. A., Yang, S. S., Miller, S. S., Bucciarelli, B., Liu, J., Rydeen, A., et al. (2013). An RNA-seq transcriptome analysis of orthophosphate-deficient white lupin reveals novel insights into phosphorus acclimation in plants. *Plant Physiol.* 161, 705–724. doi: 10.1104/pp.112.209254
- Pakdeechanuan, P., Shoji, T., and Hashimoto, T. (2012). Root-to-shoot translocation of alkaloids is dominantly suppressed in *Nicotiana glauca*. *Plant Cell Physiol.* 53, 1247–1254. doi: 10.1093/pcp/pcs065
- Pan, Q., Mustafa, N. R., Tang, K., Choi, Y. H., and Verpoorte, R. (2016). Monoterpenoid indole alkaloids biosynthesis and its regulation in *Catharanthus roseus*: a literature review from genes to metabolites. *Phytochem. Rev.* 15, 221–250. doi: 10.1007/s11101-015-9406-4
- Parra-González, L. B., Aravena-Abarzúa, G. A., Navarro-Navarro, C. S., Udall, J., Maughan, J., Peterson, L. M., et al. (2012). Yellow lupin (*Lupinus luteus* L.) transcriptome sequencing: molecular marker development and comparative studies. *BMC Genomics* 13:425. doi: 10.1186/1471-2164-13-425
- Pasquali, G., Goddijn, O. J., de Waal, A., Verpoorte, R., Schilperoort, R. A., Hoge, J. H. C., et al. (1992). Coordinated regulation of two indole alkaloid biosynthetic genes from *Catharanthus roseus* by auxin and elicitors. *Plant Mol. Biol.* 18, 1121–1131. doi: 10.1007/BF00047715
- Pauli, H. H., and Kutchan, T. M. (1998). Molecular cloning and functional heterologous expression of two alleles encoding (S)-N-methylcocaurine 3'-hydroxylase (CYP80B1), a new methyl jasmonate-inducible cytochrome P-450-dependent mono-oxygenase of benzylisoquinoline alkaloid biosynthesis. *Plant J.* 13, 793–801. doi: 10.1046/j.1365-3113X.1998.00085.x
- Petterson, D., Gladstones, J., Atkins, C., and Hamblin, J. (1998). “Composition and food uses of lupins,” in *Lupins as Crop Plants: Biology, Production and Utilization*, eds J. S. Gladstones, C. Atkins, and J. Hamblin (Cambridge: CAB International), 353–384.
- Petterson, D. S., Sipsas, S., and Mackintosh, J. B. (1997). *The Chemical Composition and Nutritive Value of Australian Pulses*. Canberra: Grains Research and Development Corporation.
- Philippi, J., Schliephake, E., Jürgens, H. U., Jansen, G., and Ordon, F. (2015). Feeding behavior of aphids on narrow-leaved lupin (*Lupinus angustifolius*) genotypes varying in the content of quinolizidine alkaloids. *Entomol. Exp. Appl.* 156, 37–51. doi: 10.1111/eea.12313
- Pigeaire, A., Abernethy, D., Smith, P. M., Simpson, K., Fletcher, N., Lu, C.-Y., et al. (1997). Transformation of a grain legume (*Lupinus angustifolius* L.) via *Agrobacterium tumefaciens*-mediated gene transfer to shoot apices. *Mol. Breed.* 3, 341–349. doi: 10.1023/a:1009642620907
- Pniewski, T., Kapusta, J., and Plucienniczak, A. (2006). *Agrobacterium*-mediated transformation of yellow lupin to generate callus tissue producing HBV surface antigen in a long-term culture. *J. Appl. Genet.* 47, 309–318. doi: 10.1007/BF03194640
- Polowick, P. L., Loukanina, N. N., and Doshi, K. M. (2014). *Agrobacterium*-mediated transformation of tarwi (*Lupinus mutabilis* Sweet), a potential platform for the production of plant-made proteins. *Vitro Cell. Dev. Biol. Plant* 50, 401–411. doi: 10.1007/s11627-014-9601-9
- Priddis, C. R. (1983). Capillary gas chromatography of lupin alkaloids. *J. Chromatogr. A* 261, 95–101. doi: 10.1016/S0021-9673(01)87922-X
- Reinhard, H., Rupp, H., Sager, F., Streule, M., and Zoller, O. (2006). Quinolizidine alkaloids and phomopsins in lupin seeds and lupin containing food. *J. Chromatogr. A* 1112, 353–360. doi: 10.1016/j.chroma.2005.11.079
- Resta, D., Boschini, G., D'Agostina, A., and Arnoldi, A. (2008). Evaluation of total quinolizidine alkaloids content in lupin flours, lupin-based ingredients, and foods. *Mol. Nutr. Food Res.* 52, 490–495. doi: 10.1002/mnfr.200700206
- Ridsdill-Smith, J., Edwards, O., Wang, S., Ghisalberti, E., Reidy-Crofts, J., Simon, J., et al. (2004). “Aphid response to plant defensive compounds in lupins,” in *Proceedings of the Sixth International Symposium on Aphids*, (Paris: Institut National de la Recherche Agronomique), 491–497.
- Rodriguez-Medina, C., Atkins, C. A., Mann, A. J., Jordan, M. E., and Smith, P. M. (2011). Macromolecular composition of phloem exudate from white lupin (*Lupinus albus* L.). *BMC Plant Biol.* 11:36–54. doi: 10.1186/1471-2229-11-36
- Roepke, J., Salim, V., Wu, M., Thamm, A. M., Murata, J., Ploss, K., et al. (2010). Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15287–15292. doi: 10.1073/pnas.0911451107
- Saito, K., Koike, Y., Suzuki, H., and Murakoshi, I. (1993). Biogenetic implication of lupin alkaloid biosynthesis in bitter and sweet forms of *Lupinus luteus* and *L. albus*. *Phytochemistry* 34, 1041–1044. doi: 10.1016/S0031-9422(00)90709-X
- Saito, K., Suzuki, H., Takamatsu, S., and Murakoshi, I. (1992). Acyltransferases for lupin alkaloids in *Lupinus hirsutus*. *Phytochemistry* 32, 87–91. doi: 10.1016/0031-9422(92)80112-R
- Samanani, N., Liscombe, D. K., and Facchini, P. J. (2004). Molecular cloning and characterization of norcocaurine synthase, an enzyme catalyzing the first committed step in benzylisoquinoline alkaloid biosynthesis. *Plant J.* 40, 302–313. doi: 10.1111/j.1365-3113X.2004.02210.x
- Sbabou, L., Bucciarelli, B., Miller, S., Liu, J., Berhada, F., Filali-Maltouf, A., et al. (2010). Molecular analysis of SCARECROW genes expressed in white lupin cluster roots. *J. Exp. Bot.* 61, 1351–1363. doi: 10.1093/jxb/erp400
- Secco, D., Shou, H., Whelan, J., and Berkowitz, O. (2014). RNA-seq analysis identifies an intricate regulatory network controlling cluster root development in white lupin. *BMC Genomics* 15:1. doi: 10.1186/1471-2164-15-230
- Shitan, N., Bazin, I., Dan, K., Obata, K., Kigawa, K., Ueda, K., et al. (2003). Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. *Proc. Natl. Acad. Sci. U.S.A.* 100, 751–756. doi: 10.1073/pnas.0134257100
- Shitan, N., Dalmas, F., Dan, K., Kato, N., Ueda, K., Sato, F., et al. (2013). Characterization of *Coptis japonica* CjABC2, an ATP-binding cassette protein involved in alkaloid transport. *Phytochemistry* 91, 109–116. doi: 10.1016/j.phytochem.2012.02.012
- Shitan, N., Kiuchi, F., Sato, F., Yazaki, K., and Yoshimatsu, K. (2005). Establishment of *Rhizobium*-mediated transformation of *Coptis japonica* and molecular analyses of transgenic plants. *Plant Biotechnol.* 22, 113–118. doi: 10.5511/plantbiotechnology.22.113
- Shitan, N., Minami, S., Morita, M., Hayashida, M., Ito, S., Takanashi, K., et al. (2014). Involvement of the leaf-specific multidrug and toxic compound extrusion (MATE) transporter Nt-JAT2 in vacuolar sequestration of nicotine in *Nicotiana tabacum*. *PLoS ONE* 9:e108789. doi: 10.1371/journal.pone.0108789
- Shoji, T., Inai, K., Yazaki, Y., Sato, Y., Takase, H., Shitan, N., et al. (2009). Multidrug and toxic compound extrusion-type transporters implicated in vacuolar sequestration of nicotine in tobacco roots. *Plant Physiol.* 149, 708–718. doi: 10.1104/pp.108.132811
- Shoji, T., Kajikawa, M., and Hashimoto, T. (2010). Clustered transcription factor genes regulate nicotine biosynthesis in tobacco. *Plant Cell* 22, 3390–3409. doi: 10.1105/tpc.110.078543
- Shoji, T., Ogawa, T., and Hashimoto, T. (2008). Jasmonate-induced nicotine formation in tobacco is mediated by tobacco COI1 and JAZ genes. *Plant Cell Physiol.* 49, 1003–1012. doi: 10.1093/pcp/pcn077
- Suttipanta, N., Pattanaik, S., Kulshrestha, M., Patra, B., Singh, S. K., and Yuan, L. (2011). The transcription factor CrWRKY1 positively regulates the terpenoid indole alkaloid biosynthesis in *Catharanthus roseus*. *Plant Physiol.* 157, 2081–2093. doi: 10.1104/pp.111.181834
- Suzuki, H., Koike, Y., Murakoshi, I., and Saito, K. (1996). Subcellular localization of acyltransferases for quinolizidine alkaloid biosynthesis in *Lupinus*. *Phytochemistry* 42, 1557–1562. doi: 10.1016/0031-9422(96)00195-1
- Suzuki, H., Murakoshi, I., and Saito, K. (1994). A novel O-tigloyltransferase for alkaloid biosynthesis in plants. Purification, characterization, and distribution in *Lupinus* plants. *J. Biol. Chem.* 269, 15853–15860.
- Sweetingham, M., Kingwell, R., Palta, J., and Berger, J. B. (2008). “Lupins—reflections and future possibilities,” in *Proceedings of the 12th International Lupin Conference* (Geraldton: International Lupin Association), 514–524.
- Szabó, B., Tyihák, E., Szabó, G., and Botz, L. (2003). Mycotoxin and drought stress induced change of alkaloid content of *Papaver somniferum* plantlets. *Acta Bot. Hung.* 45, 409–417. doi: 10.1556/ABot.45.2003.3-4.15
- Tabé, L., Wirtz, M., Molvig, L., Droux, M., and Hell, R. (2010). Overexpression of serine acetyltransferase produced large increases in O-acetylserine and free

- cysteine in developing seeds of a grain legume. *J. Exp. Bot.* 61, 721–733. doi: 10.1093/jxb/erp338
- Thamm, A. M., Qu, Y., and De Luca, V. (2016). Discovery and metabolic engineering of iridoid/secoiridoid and monoterpenoid indole alkaloid biosynthesis. *Phytochem. Rev.* 15, 339–361. doi: 10.1007/s11101-016-9468-y
- Todd, A. T., Liu, E., Polvi, S. L., Pammett, R. T., and Page, J. E. (2010). A functional genomics screen identifies diverse transcription factors that regulate alkaloid biosynthesis in *Nicotiana benthamiana*. *Plant J.* 62, 589–600. doi: 10.1111/j.1365-3113.2010.04186.x
- Uhde-Stone, C., Liu, J., Zinn, K. E., Allan, D. L., and Vance, C. P. (2005). Transgenic proteoid roots of white lupin: a vehicle for characterizing and silencing root genes involved in adaptation to P stress. *Plant J.* 44, 840–853. doi: 10.1111/j.1365-3113.2005.02573.x
- Van Moerkercke, A., Steensma, P., Schweizer, F., Pollier, J., Gariboldi, I., Payne, R., et al. (2015). The bHLH transcription factor BIS1 controls the iridoid branch of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8130–8135. doi: 10.1073/pnas.1504951112
- von Baer, E., and Gross, R. (1977). Selection of low alkaloid forms of *Lupinus mutabilis*. *Z. Pflanzenzüchtung* 79, 52–58.
- von Baer, E., and Gross, R. (1983). Sweet strains of *Lupinus mutabilis*. *Z. Pflanzenzüchtung* 91, 334–337.
- von Sengbusch, R. (1942). Sweet lupins and oil lupins. The history of the origin of some new crop plants. *Landwirtsch. Jahrb.* 91, 719–880.
- Waller, G. R., and Nowacki, E. K. (1978). *Alkaloid Biology and Metabolism in Plants*. New York, NY: Plenum Press, 129–133.
- Wang, S., Liu, A., Ridsdill-Smith, T., and Ghisalberti, E. (2000). Role of alkaloids in resistance of yellow lupin to red-legged earth mite *Halotydeus destructor*. *J. Chem. Ecol.* 26, 429–441. doi: 10.1023/A:1005413606680
- Wang, X., and Bennetzen, J. L. (2015). Current status and prospects for the study of *Nicotiana* genomics, genetics, and nicotine biosynthesis genes. *Mol. Genet. Genomics* 290, 11–21. doi: 10.1007/s00438-015-0989-7
- Wang, Z., Straub, D., Yang, H., Kania, A., Shen, J., Ludewig, U., et al. (2014). The regulatory network of cluster-root function and development in phosphate-deficient white lupin (*Lupinus albus*) identified by transcriptome sequencing. *Physiol. Plant.* 151, 323–338. doi: 10.1111/ppl.12187
- Wijayanto, T., Barker, S. J., Wylie, S. J., Gilchrist, D. G., and Cowling, W. A. (2009). Significant reduction of fungal disease symptoms in transgenic lupin (*Lupinus angustifolius*) expressing the anti-apoptotic baculovirus gene p35. *Plant Biotechnol. J.* 7, 778–790. doi: 10.1111/j.1467-7652.2009.00443.x
- Wink, M. (1983). Wounding-induced increase of quinolizidine alkaloid accumulation in lupin leaves. *Z. Naturforsch. C* 38, 905–909. doi: 10.1515/znc-1983-11-1204
- Wink, M. (1987a). Quinolizidine alkaloids: biochemistry, metabolism, and function in plants and cell suspension cultures. *Planta Med.* 53, 509–514. doi: 10.1055/s-2006-962797
- Wink, M. (1987b). Site of lupanine and sparteine biosynthesis in intact plants and in vitro organ cultures. *Z. Naturforsch. C* 42, 868–872. doi: 10.1515/znc-1987-7-823
- Wink, M. (1990). “Plant breeding: low or high alkaloid content,” in *Proceedings of the 6th International Lupin Conference* (Geraldton: International Lupin Association), 326–334.
- Wink, M. (1992). The role of quinolizidine alkaloids in plant-insect interactions. *Insect Plant Interact.* 4, 131–166.
- Wink, M. (1993). “Quinolizidine alkaloids,” in *Methods in Plant Biochemistry*, ed. P. Waterman (London: Academic Press), 197–239.
- Wink, M. (1994). “Biological activities and potential application of lupin alkaloids,” in *Advances in Lupin Research*, eds J. M. Neves-Martins and M. L. Beirao da Costa (Lisbon: ISA Press), 161–178.
- Wink, M., and Hartmann, T. (1981). Activation of chloroplast-localized enzymes of quinolizidine alkaloid biosynthesis by reduced thioredoxin. *Plant Cell Rep.* 1, 6–9. doi: 10.1007/BF00267646
- Wink, M., and Hartmann, T. (1982a). Enzymatic synthesis of quinolizidine alkaloid esters: a tigloyl-CoA: 13-hydroxylupanine O-tigloyltransferase from *Lupinus albus* L. *Planta* 156, 560–565. doi: 10.1007/BF00392781
- Wink, M., and Hartmann, T. (1982b). Localization of the enzymes of quinolizidine alkaloid biosynthesis in leaf chloroplasts of *Lupinus polyphyllus*. *Plant Physiol.* 70, 74–77. doi: 10.1104/pp.70.1.74
- Wink, M., Meißner, C., and Witte, L. (1995). Patterns of quinolizidine alkaloids in 56 species of the genus *Lupinus*. *Phytochemistry* 38, 139–153. doi: 10.1016/0031-9422(95)91890-D
- Wink, M., and Witte, L. (1984). Turnover and transport of quinolizidine alkaloids. Diurnal fluctuations of lupanine in the phloem sap, leaves and fruits of *Lupinus albus* L. *Planta* 161, 519–524. doi: 10.1007/BF00407083
- Yamada, Y., Koyama, T., and Sato, F. (2011). Basic helix-loop-helix transcription factors and regulation of alkaloid biosynthesis. *Plant Signal. Behav.* 6, 1627–1630. doi: 10.4161/psb.6.11.17599
- Yamagishi, M., Masuta, C., Suzuki, M., and Netsu, O. (2015). Peanut stunt virus-induced gene silencing in white lupin (*Lupinus albus*). *Plant Biotechnol.* 32, 181–191. doi: 10.5511/plantbiotechnology.15.0521a
- Yang, H., Tao, Y., Zheng, Z., Zhang, Q., Zhou, G., Sweetingham, M. W., et al. (2013). Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS ONE* 8:e64799. doi: 10.1371/journal.pone.0064799
- Yazaki, K., Sugiyama, A., Morita, M., and Shitan, N. (2008). Secondary transport as an efficient membrane transport mechanism for plant secondary metabolites. *Phytochem. Rev.* 7, 513–524. doi: 10.1007/s11101-007-9079-8
- Yu, F., and De Luca, V. (2013). ATP-binding cassette transporter controls leaf surface secretion of anticancer drug components in *Catharanthus roseus*. *Proc. Natl. Acad. Sci. U.S.A.* 110, 15830–15835. doi: 10.1073/pnas.1307504110
- Zachow, F. (1967). A new gene for low alkaloid content in *Lupinus angustifolius*. *Der Züchter* 37, 35–38. doi: 10.1007/BF00621153
- Zhang, H., Hedhili, S., Montiel, G., Zhang, Y., Chatel, G., Pré, M., et al. (2011). The basic helix-loop-helix transcription factor CrMYC2 controls the jasmonate-responsive expression of the ORCA genes that regulate alkaloid biosynthesis in *Catharanthus roseus*. *Plant J.* 67, 61–71. doi: 10.1111/j.1365-3113.2011.04575.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Frick, Kamphuis, Siddique, Singh and Foley. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Transcriptome Analysis of a New Peanut Seed Coat Mutant for the Physiological Regulatory Mechanism Involved in Seed Coat Cracking and Pigmentation

Liyun Wan¹, Bei Li¹, Manish K. Pandey², Yanshan Wu¹, Yong Lei¹, Liying Yan¹, Xiaofeng Dai³, Huifang Jiang¹, Juncheng Zhang¹, Guo Wei³, Rajeev K. Varshney^{2,4} and Boshou Liao^{1*}

¹ Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of Chinese Academy of Agricultural Sciences, Wuhan, China, ² Center of Excellence in Genomics, International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India, ³ Institute of Food Science and Technology of Chinese Academy of Agricultural Sciences, Beijing, China, ⁴ School of Plant Biology and Institute of Agriculture, The University of Western Australia, Crawley, WA, Australia

OPEN ACCESS

Edited by:

Nicolas Rispaï,
Institute for Sustainable Agriculture
(CSIC), Spain

Reviewed by:

Xingjun Wang,
Shandong Academy of Agricultural
Sciences, China
Haiwen Zhang,
Biotechnology Research Institute
(CAAS), China

*Correspondence:

Boshou Liao
lboshou@hotmail.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 12 July 2016

Accepted: 20 September 2016

Published: 14 October 2016

Citation:

Wan L, Li B, Pandey MK, Wu Y, Lei Y,
Yan L, Dai X, Jiang H, Zhang J, Wei G,
Varshney RK and Liao B (2016)
Transcriptome Analysis of a New
Peanut Seed Coat Mutant for the
Physiological Regulatory Mechanism
Involved in Seed Coat Cracking and
Pigmentation.
Front. Plant Sci. 7:1491.
doi: 10.3389/fpls.2016.01491

Seed-coat cracking and undesirable color of seed coat highly affects external appearance and commercial value of peanuts (*Arachis hypogaea* L.). With an objective to find genetic solution to the above problems, a peanut mutant with cracking and brown colored seed coat (testa) was identified from an EMS treated mutant population and designated as “peanut seed coat crack and brown color mutant line (*pscb*).” The seed coat weight of the mutant was almost twice of the wild type, and the germination time was significantly shorter than wild type. Further, the mutant had lower level of lignin, anthocyanin, proanthocyanidin content, and highly increased level of melanin content as compared to wild type. Using RNA-Seq, we examined the seed coat transcriptome in three stages of seed development in the wild type and the *pscb* mutant. The RNA-Seq analysis revealed presence of highly differentially expressed phenylpropanoid and flavonoid pathway genes in all the three seed development stages, especially at 40 days after flowering (DAF40). Also, the expression of polyphenol oxidases and peroxidase were found to be activated significantly especially in the late seed developmental stage. The genome-wide comparative study of the expression profiles revealed 62 differentially expressed genes common across all the three stages. By analyzing the expression patterns and the sequences of the common differentially expressed genes of the three stages, three candidate genes namely *c36498_g1* (*CCoAOMT1*), *c40902_g2* (*kinesin*), and *c33560_g1* (*MYB3*) were identified responsible for seed-coat cracking and brown color phenotype. Therefore, this study not only provided candidate genes but also provided greater insights and molecular genetic control of peanut seed-coat cracking and color variation. The information generated in this study will facilitate further identification of causal gene and diagnostic markers for breeding improved peanut varieties with smooth and desirable seed coat color.

Keywords: peanut (*Arachis hypogaea*), seed-coat cracking, pigmentation, RNA-seq, flavonoid pathway

INTRODUCTION

A typical peanut (*Arachis hypogaea* L.) seed has three parts i.e., seed coat (also known as testa), embryo and endosperm. Seed coat is the outer protective layer of seed and one of its major role is to provide protection to embryo and endosperm from external factors such as infection of insects, bacteria, fungi and virus, mechanical injuries, and even desiccation of the seed. In legumes including peanut, the seed coat, and endosperm develop first, followed by the embryo (Weber et al., 2005). Rapid cotyledon growth sometimes may not adequately match the expansion of the seed coat leading to formation of cracks in seed coat (Agarwal and Menon, 1974). In other words, the seed-coat cracking (SC) results from the separation of epidermal (palisade cells) and hypodermal tissues leading to exposure of the underlying parenchyma tissues (Wolf and Baker, 1972). The most adverse effect of SC is that seeds become more vulnerable to storage problems and field microorganisms, leading to seed rotting or pre- and post-emergence damping under high humid conditions. Although the reason for the physical separation of palisade and hypodermal cells is not well-known, genetic and environmental factors have been implicated for SC in other crops such as soybean (Stewart and Wentz, 1930; Woodworth and Williams, 1938; Liu, 1949; Schlub and Schmitthenner, 1978; Duke et al., 1983, 1986) and watermelon (Hafez et al., 1981). The seed coat cracks in soybean were linked to physiological and ultimate structure of the cell wall (Kour et al., 2014).

Lignin is a complex and heterogeneous polymer that constitutes one of the major components of the secondary wall of xylem cells and fibers (Mellerowicz et al., 2001). Lignification confers not only the mechanical support and optimizes transport of water and solutes along vascular system but also protects against pathogens (Boerjan et al., 2003). Lignin is the second most abundant biological product in nature, and is formed by oxidative polymerization of three main constituents, namely monolignols p-coumaryl, coniferyl, and sinapyl alcohols through the phenylpropanoid pathway. Once incorporated in the lignin polymer, these precursors are known as p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) subunits, respectively (Anterola and Lewis, 2002; Boerjan et al., 2003). Lignins together with anthocyanins, flavonols and proanthocyanidins constitute the main group of plant phenylpropanoids (Fornalé et al., 2010).

The seed coat color varies in different species and genotypes and even at different seed developmental stages. Flavonoids are the major secondary metabolites influencing seed coat color in plants and represent a highly diverse group of plant aromatic secondary metabolites. The major forms of flavonoids include anthocyanins (red to purple pigments), flavonols (colorless to pale yellow pigments), and proanthocyanidins (PAs), also known as condensed tannins (colorless pigments that brown with oxidation). These flavonoids are present in varied proportions and quantity in different plant species, organs, developmental stages and environmental conditions. The PAs, oligomers of flavan-3-ol units, have received particular attention due to their abundance in seed coats (Dixon et al., 2005). The mechanism of seed coat pigmentation was well-studied in model plants such as *Arabidopsis* wherein the flavonols and proanthocyanidins

derivatives were found responsible for the pigmentation pattern of seeds in addition to their involvement in a wide range of biological functions (Shirley et al., 1995). In the mature testa, flavonoids were detected in both the endothelium and the three crushed parenchymal layers just above the endothelium (Debeaujon et al., 2000). PAs have been shown to accumulate exclusively in the endothelium layer (Devic et al., 1999).

In recent years, several efforts were made for elucidating the flavonoid biosynthetic pathway from the molecular genetics point of view (Winkel-Shirley, 2001; Tanaka et al., 2008). Mutants affecting flavonoid synthesis were isolated in a variety of plant species based on alterations in flower and seed pigmentation. Recent studies conducted in *Arabidopsis* helped in developing some understanding on regulation and subcellular organization of the flavonoid pathway (Winkel-Shirley, 2001). The same study also indicated that the genetic loci for both structural and regulatory genes were scattered across the *Arabidopsis* genome and were identified largely on the basis of mutations that abolish or reduce pigmentation in the seed coat. The major functional and regulatory genes in flavonoid metabolism include *PAL*, *C4H*, *4CL*, *CHS*(*TT4*), *CHI*(*TT5*), *F3H*(*TT6*), *F3'H*(*TT7*), *DRF*(*TT3*), *ANS*(*LODX/TT18*), *LAR/LCR*, *BAN*(*ANR*), *TT12*, *TT19*(*GST*), *TT10*, *FLS*, and *AHA10* etc. (Chapple et al., 1994; Winkel-Shirley, 2001; Abrahams et al., 2003). Regulatory proteins controlling flavonoid biosynthesis were also characterized e.g., MYB-bHLH-WDR (MBW) complex was found to be involved in biosynthesis of PAs and anthocyanins (Baudry et al., 2006; Lepiniec et al., 2006) and the R2R3-MYBs PRODUCTION OF FLAVONOL GLYCOSIDE (PFG1/MYB12, PFG2/MYB11, and PFG3/MYB111) positively regulated flavonol biosynthesis in root and the aerial parts (Dubos et al., 2010; Stracke et al., 2010a,b), whereas single repeat small MYBs CAPRICE (CPC) or MYBL2 was found to be involved in negatively regulating anthocyanin synthesis (Dubos et al., 2008; Zhu et al., 2009).

In peanut, previous studies mainly focused on identifying the antioxidant of seed coat and the extraction pigments (Wang et al., 2007; Ballard et al., 2009; Zhang et al., 2013; de Camargo et al., 2014; Ma et al., 2014). These studies showed that peanut seed coat with different colors were composed of different pigment composition. However, none of the above mentioned studies provided any information on peanut seed-coat cracking and pigmentation mechanism. In this study, we first identified a spontaneous seed coat-cracking and seed color mutant from Zhonghua16, designated as "*pscb*," and then employed RNA-seq approach to develop better understanding of the mechanism of seed coat-cracking and brown color development in seed coat of peanut.

MATERIALS AND METHODS

Plant Materials and RNA Isolation

The seed coat crack and brown testa mutant *pscb* was isolated from an ethyl methanesulfonate (EMS)-mutated population originated from an improved peanut cultivar, Zhonghua 16, with high yield and high oil content. All plants were planted in the experimental farm at the Oil Crops Research Institute (OCRI) in Wuhan.

The wild type (WT) and *pscb* mutant (M6 generation) were planted in the same field (Wuhan, China). Seed coat samples were taken at 20, 40, and 60 days after flowering (DAF) from 10 different plants in 2014. Twelve representative seeds were sampled from each seedling at each developmental stage of both the wild type and the mutant. Three biological replicates were designed. The testa separated from the sample seeds was sliced. The sliced WT and the *pscb* mutant testa samples were then frozen rapidly in liquid nitrogen and kept at -80°C , and were later used for extracting the RNA using the Tiangen RNA extraction kit (category number DP432).

Seed Water Uptake and Germination Assays

Seeds used for permeability study were harvested in 2014. For the WT and *pscb* materials, 30 seeds were tested with three replicates. The seeds were weighed, immersed in tap water for each specific time, removed from the water, blotted with cellulose tissue, weighed again, and kept again into the water. Seeds were weighed at the intervals of 30-min and 60-min during the first 8 h; at 60-min intervals during the last 6 h; and a final measurement at 24 h. The rate of water uptake was calculated by expressing it as weight increase (g) per gram seed weight (initial).

For the seed germination test, seeds were incubated in petri dishes (9 cm diameter) over two layers of medium-speed qualitative filter papers. A total of 20 seeds were placed in each petri dish and added 12 ml of sterile water. Complete experiment was performed in three replications. The seeds were incubated in a 25°C incubator with darkness. Germination was determined based on the radicle breaking through the seed coat. The germination percentage was calculated and recorded at different time points.

Tissue Preparation and Light Microscopy Observations

Peanut seeds were harvested at 20, 30, 40, 50, and 60 DAF and immediately were fixed for 24 h at 4°C in a fixation solution containing 5% acetic acid, 5% formaldehyde, and 50% ethanol. Following fixation, seeds were dehydrated at 60 min intervals through a 20% step-graded series of ethanol-water mixtures, and ended at 100% ethanol. Then, the seeds were processed at 60 min intervals through a 30% step-graded series of ethanol-TBA (tert-butyl alcohol) mixtures, and ended at 100% TBA. Seeds were subsequently infiltrated over a 24 h period with saturated paraffin-TBA mixtures, and then embedded for 48 h period in paraffin. Blocks were completely polymerized at 4°C . Semi-thin ($5\text{--}8\text{ }\mu\text{m}$ thick) sections were cut with a microtome blade KD-P (Zhejiang Jinhua Kedi Instrumental Equipment CO., LTD, China) and viewed under a stereo microscope (SZX12, Olympus, Japan). Sections were stained with TBO and observed with a Nikon ECLIPSE TI-SR microscope (Nikon Instruments, Japan).

Quantification of Lignin, Anthocyanin, Proanthocyanidin, and Phytomelanin Content

The lignin content was analyzed following Kirk and Obst (1988) and Hoebler et al. (1989), and the extraction of anthocyanins

was performed as per Pang et al. (2009). For PA analysis, 0.5–0.75 g of ground samples were extracted using extraction solution containing 5 mL of 70% acetone/0.5% acetic acid. The samples were vortexed and then sonicated at room temperature for 1 h. Following centrifugation at 2500 g for 10 min, the residues were re-extracted twice following the same above mentioned procedure. The pooled supernatants were then extracted three times using chloroform, once with hexane. The supernatants (containing soluble PAs) and residues (containing insoluble PAs) from each sample were freeze dried separately and were then suspended in extraction solution. Total soluble PA content was determined using Spectrophotometer after reaction with DMACA reagent (0.2% [w/v] DMACA in methanol-3 N HCl) at 640 nm, with (+)-catechin as standard.

For quantification of insoluble PAs, 2 mL of butanol-HCl (95:5, v/v) was added to the dried residues and the mixtures were sonicated at room temperature for 1 h, followed by centrifugation at 2500 g for 10 min. The absorption of the supernatants was measured at 550 nm. The samples were then boiled for 1 h and cooled to room temperature, and were measured again. The values were recorded by subtracting the first value from the second. Absorbance values were converted into PA equivalents using a standard curve generated with procyanidin B1 (Indofine). The extraction and characterization of phytomelanins of mature seeds were described in detail by Park et al. (2007). The phytomelanin pigments were extracted from 1 g seeds in 5 mL of 0.5 M NaOH for 1 h. The extracts were purified, and were diluted 20 times with 0.5 M NaOH, and the dilutions were subjected to absorbance measurement at 280 nm using an Ultrospec EPOCH (BioTek, China).

RNA-Seq, Data Processing, and Gene Annotation

According to the mutant seed coat color and crack phenotype, the WT and *pscb* mutant seed coat were harvested at DAF20, DAF40, and DAF60 followed by RNA sequencing using Illumina HiSeqTM2500 platform at the Novogene company (Beijing) in 2014. Briefly, $3\text{ }\mu\text{g}$ of the total RNA of each sample was used to enrich the mRNA and to construct cDNA libraries. High quality reads (clean reads) were obtained by removing low-quality reads with ambiguous nucleotides, and cutting the adaptor sequences. Transcripts were assembled using Trinity (Grabherr et al., 2011) while gene expression levels were calculated using RPKM (reads per kb per million reads) method of RSEM (Li and Dewey, 2011). Gene function was annotated based on multiple databases namely Nr (NCBI non-redundant protein sequences), Nt (NCBI non-redundant nucleotide sequences), Pfam (Protein family), KOG/COG (Clusters of Orthologous Groups of proteins), Swiss-Prot (A manually annotated and reviewed protein sequence database), KO (KEGG Ortholog database), and GO (Gene Ontology). The GO enrichment analysis of the differentially expressed genes (DEGs) was implemented by the Goseq R packages based Wallenius non-central hyper-geometric distribution (Young et al., 2010), which can adjust for gene length bias in DEGs. KOBAS (Mao et al., 2005) was used to perform KEGG pathway enrichment for the differential expression genes. Picard-tools (v1.41) and samtools

(v0.1.18) were used to sort, remove duplicated reads and merge the bam alignment results of each sample.

qRT-PCR Analysis

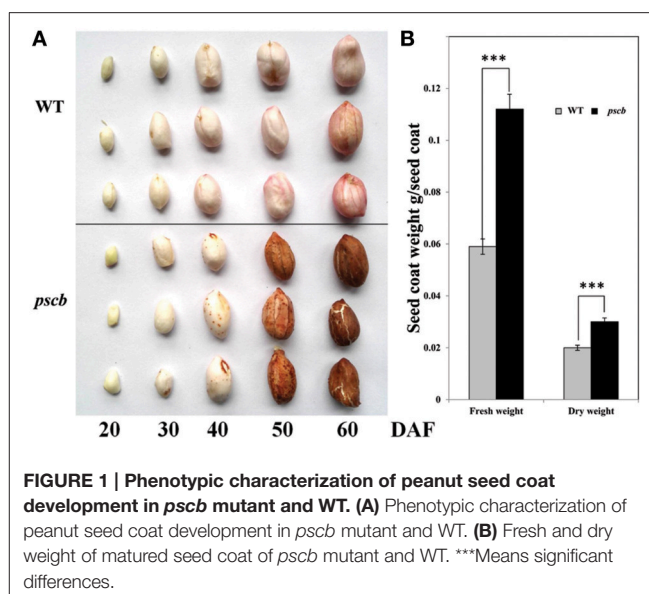
The reverse transcriptions were performed using an Invitrogen SuperScript Reagent Kit. The primer was designed using the Oligo6 software. For RT-PCR, the SYBR[®] Premix ExTaq[™] (TAKARA) was used on a Bio-Rad IQ5 real-time PCR detection system (Bio-Rad, Hercules, CA). Gene expression was analyzed for samples at 20, 40, and 60 DAF of WT and mutant. All reactions for each gene were performed in triplicate. The relative expression level of each gene among samples was calculated using the $2^{-\Delta\Delta C_t}$ method with normalization to the internal reference actin gene. The parameters of thermal cycle were 95°C for 30 s, followed by 40 cycles of 95°C for 10 s and 50–56°C for 25 s at a volume of 20 μ l.

RESULTS

Phenotypic Variation between Seed Coat of Wild and Mutant Genotypes

A peanut mutant with cracked and brown color seed coat named *pscb* was identified from a mutant population treated with 1.0% EMS in the background of an elite peanut cultivar Zhonghua16 (WT). In M3 generation, the ratio of *pscb* mutant and the normal plants was 1:3 (69 *pscb* and 206 normal). Although there was no difference in the seed coat of WT and the *pscb* mutant at the early stage, however, few tiny brown points appeared in the seed coat of the *pscb* mutant at the stage of DAF40 when the seed coat of WT turned pink gradually. Interestingly, the seed coat of *pscb* turned totally brown while the WT were still in pink at DAF60. It was observed that the seed coat developed cracks when plants reached their physiological maturity and seeds were fully developed. At stage DAF60, seed coat cracks became more evident and wide (Figure 1). It is important to mention that cracks only appeared in the outer layer of the seed coat and not in the inner integument (Figure 1). The seed coat cracks in mutant appeared in all the growing conditions (3 years in Wuhan and 3 years in Zhanjiang) in varied intensities i.e., from a minute or invisible crack to several wide cracks. We observed that the mature seed coat of *pscb* mutant was much thicker than the WT. In addition, the seed coat of *pscb* mutant had 190% higher fresh weight and 150% higher dry weight than that of WT.

The seed coat (testa) of higher plants protects the embryo against adverse environmental conditions including germination control through dormancy imposition and by limiting the detrimental activity of physical and biological agents during seed storage. We subsequently speculated that there may be significant differences between WT and *pscb* mutant in seed water uptake and germination. To test this hypothesis, the water uptake and germination experiments were carried out in the WT and the *pscb* mutant. After keeping seeds immersed in water for 0.5 h, the *pscb* mutant absorbed 11.19% of the seed dry weight water while it was only 7.87% in wild type. The water absorption in *pscb* mutant was recorded 14.75, 17.69, 21.16, 24.37, 27.75, 38.94, 42.32, 44.06, 54.12, 54.12, and 54.12% at time interval of 1, 1.5, 2, 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, 18.5, 19.5, and 20.5 h, respectively. While at each paired time points, the water absorbed by the wild



type was 1.58–6.43% lower than that in *pscb* mutant (Figure 2A). After absorption of enough water, the seeds of *pscb* mutant and wild type were transferred to a 25°C incubator. The germination rate of wild type was 29.40% as compared to only 15.4% observed in *pscb* mutant, the differences enlarged till 24 h and 36 h and then narrowed at 48 h. Till 60 h, all the seeds germinated (Figure 2B) and only the length of bacon differed between wild type and *pscb* mutant (Figure 2C). In other words, the results showed that the *pscb* mutant had faster water uptake efficiency and delayed germination compared to WT.

Estimation of Lignin, Anthocyanin, and Proanthocyanidin Contents

It is well-known fact that the seed crack is related to physiological and ultimate structure of the cell wall. Since the lignin in the early stages of peanut seed coat is extremely low in quantity to detect, we analyzed the lignin content of the mature seed coat of the *pscb* mutant and the WT. The results showed that the lignin content in *pscb* mutant was 40.2% less as compared to WT (from 19.38 to 7.79 mg/g FW; Figure 3A). The *pscb* mutant showed brown seed coat color from DAF40, indicating change in the seed coat pigments. Direct measurement of anthocyanins and PAs content confirmed significant difference between *pscb* mutant and WT seed coat extracts (Figures 3B,D). The anthocyanins content was obviously lower i.e., 9.9 μ g/g FW in *pscb* mutant as compared to WT i.e., 231 μ g/g FW (Figure 3B). To quantify PAs accumulation, we extracted soluble and insoluble (non-extractable) PAs separately. Soluble PAs content in WT was 0.415 μ g/g FW whereas the level of soluble PAs in *pscb* mutant was almost undetectable i.e., mere 0.0076 μ g/g FW (Figure 3C). Measurement of insoluble PAs based on butanol-HCl hydrolysis showed that *pscb* mutant had less non-extractable PAs (Figure 3D).

To determine the cellular distribution of polyphenol compounds in developing seeds of *pscb* mutant and WT seed coat, a histochemical analysis was performed in the seed coat harvested at different developmental stages. TBO

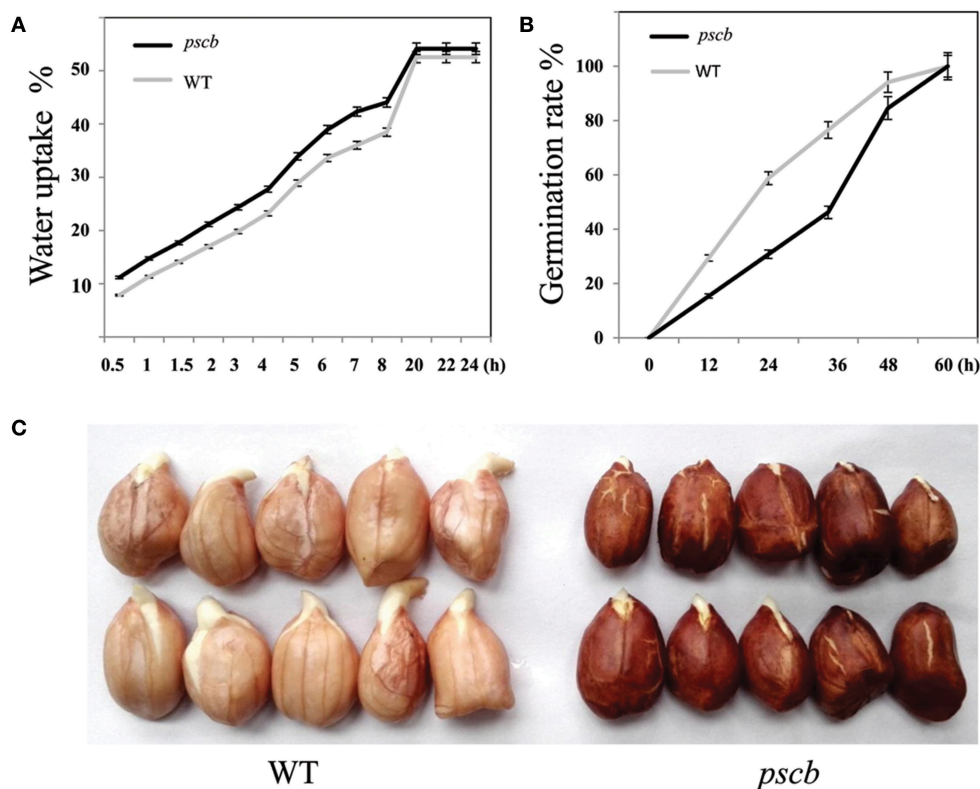


FIGURE 2 | Water uptake efficiency and germination in *pscb* mutant and WT. (A) Water uptake efficiency of *pscb* mutant and WT. **(B)** Germination time of *pscb* mutant and WT. **(C)** The radicals elongated faster in WT when compared with *pscb* mutant.

(toluidine blue O) staining of transverse sections of developing seeds revealed that at early seed development stages (DAF20 and DAF30), staining was similar in *pscb* mutant and WT (Figures 3e-1, e-2, f-1, f-2). At 40DAF, WT stained more intensely than the *pscb* mutant in the outer layer of the testa, indicating that more polymeric phenolic compounds had accumulated in WT (Figures 3e-3, f-3). The distribution of polymeric phenolic compounds was significantly different at the two late seed development stages (Figures 3e-4, e-5, f-4, f-5).

Seed Coat Transcriptome Differences between *pscb* Mutant and WT during Seed Coat Development

In order to understand the mechanism of seed coat development, we selected RNA samples of seed coat from three seed developmental stages i.e., DAF20, DAF40, and DAF60 showing different seed coat phenotype to perform the transcriptome analysis. The sequence data was deposited in the BioProject database of National Center for Biotechnology Information under the accession number PRJNA324725. One of the primary goals of transcriptome sequencing was to compare the gene expression levels in *pscb* mutant and WT. In this study, we used a stringent value of $FDR \leq 0.001$ and fold change ≥ 2 as the threshold to judge the significant differences in the gene expression. A total of 5726 genes were found differently expressed

between *pscb* mutant and WT (Table S1). Of these, 255 genes expressed at DAF20, 5443 genes at DAF40, and 341 genes at DAF60 (Figure 4). At DAF20, the number of up-regulated genes was more than the down-regulated genes (165:90). At DAF40, there was not much difference in the number of up-regulated and the down-regulated genes (2544:2899). At DAF60, the number of up-regulated genes was half of the number of the down-regulated genes (110:231). It was important to note that the number of DEGs increased significantly at the stage of DAF40, which was the vigorous growth stage during seed development (Figure 4).

To gain insights into the functional categories that were altered between *pscb* and WT (Zhonghua16), GO categories were assigned to the DEGs. Further, GO enrichment analysis of the DEGs in different developmental stages between the *pscb* mutant and WT (Zhonghua16) was performed for different developmental stages. Interestingly, no GO terms were enriched at DAF20, however, several significantly enriched terms in the biological process, molecular function, and cellular component categories were identified at DAF40 and DAF60 (Table 1). ADP binding (GO: 0043531), structural constituent of cell wall (GO: 0005199), plant-type cell wall organization or biogenesis (GO: 0071669), plant-type cell wall organization (GO: 0009664), and external encapsulating structure organization (GO: 0045229) were dominant terms at DAF60 in comparisons to DAF40. At DAF40, seed coat gets brown phenotype in

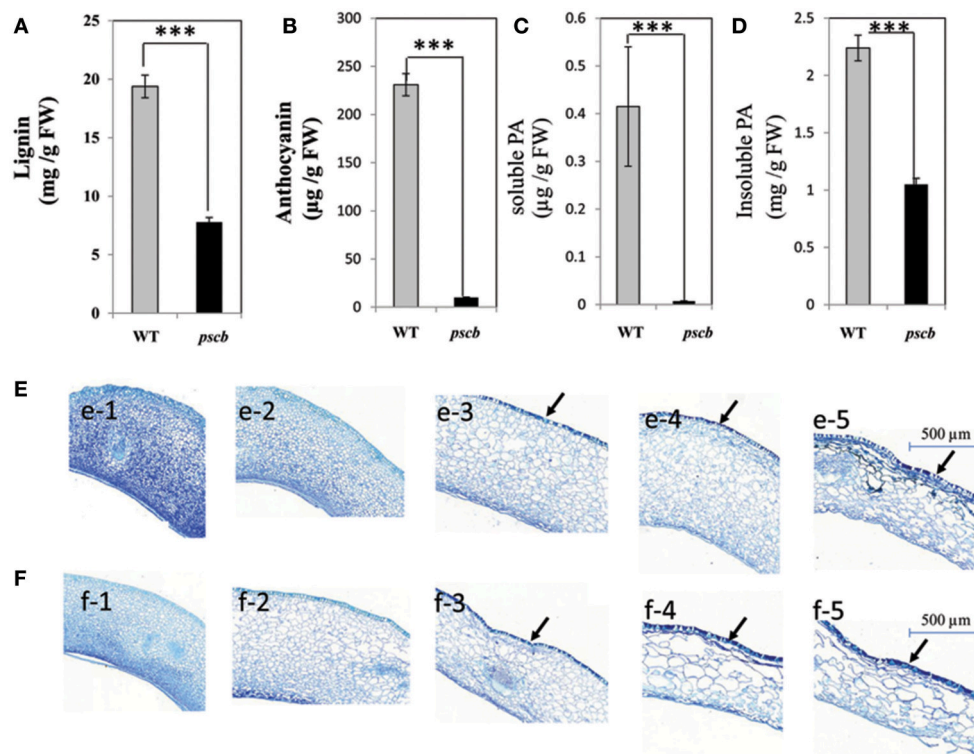


FIGURE 3 | The *pscb* mutant has decreased lignin, anthocyanin, and Proanthocyanidin content. (A) Lignin content of matured seed coat in *pscb* mutant and WT. **(B)** Anthocyanin levels from *pscb* mutant and WT matured seed coat. **(C)** Soluble PA levels of *pscb* mutant and WT seed coat. **(D)** In-soluble PA content from *pscb* mutant and WT matured seed coat. FW, Fresh weight. **(E,F)** Changes of proanthocyanidins and phenolic compounds during seed coat development of *pscb* mutant and WT. **(e1–e5)** Detection and localization of proanthocyanidin and phenolic compounds in seed coat development in *pscb* mutant. Bars = 500 μ m. **(f1–f5)** Detection and localization of proanthocyanidin and phenolic compounds in seed coat development in WT. Bars = 500 μ m. Black arrows, accumulation site for polymeric phenolic compounds. ***Means significant differences.

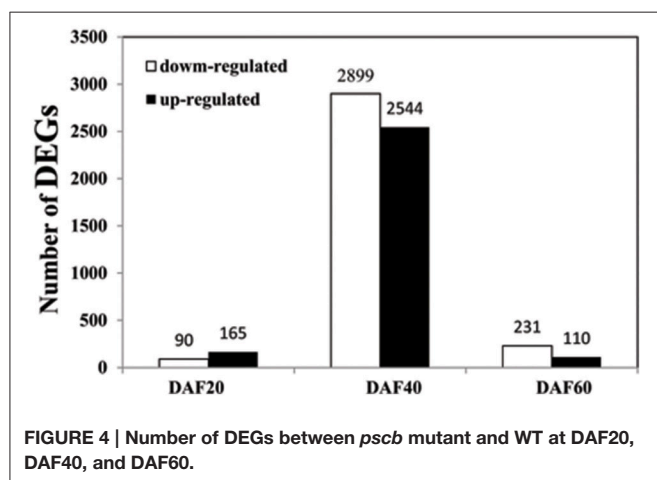


FIGURE 4 | Number of DEGs between *pscb* mutant and WT at DAF20, DAF40, and DAF60.

the *pscb* mutant, indicating the difference between cell wall organizations might play an important role in the phenotype differentiation. At the DAF60, the main GO terms were related to fatty acid synthase and oxidoreductase activity (Table 1), including 3-oxoacyl-[acyl-carrier-protein] synthase activity (GO: 0004315), fatty acid synthase activity (GO: 0004312),

fatty acid synthase complex (GO: 0005835), transferase activity, transferring acyl groups other than amino-acyl groups (GO: 0016747) etc. KEGG pathway analysis assigned the differential genes to 37, 287, and 46 metabolic pathways in three different developmental stages of *pscb* and Zhonghua 16. The complete list of metabolic pathways is provided in Table S2. Table 2 lists the metabolic/biological pathways in common of *pscb* compared with WT. Notably, among the 16 common KEGGs, 6 were involved in Tyrosine, Tryptophan, and Phenylalanine metabolism (Table 2).

Verification of Differentially Expressed Genes during Seed Coat Development in *pscb* Mutant and WT

Transcriptional regulation revealed by RNA-seq data was confirmed in a biologically independent experiment using the quantitative reverse transcription PCR. A total of 17 genes related to cell wall organization were selected to design gene-specific primers (Table S3) for real time PCR analysis (Figure 5). A linear regression analysis showed an overall correlation coefficient of $R = 0.622$, which indicated a good correlation between transcript abundance assayed by real-time PCR and the transcription profile revealed by RNA-seq data (Figure 5B).

TABLE 1 | Functional categorization of genes with significant transcriptional differences between *pscb* mutant and WT.

	GO accession	Description	Term type*	Corrected <i>p</i> -value	DEG item
DAF40	GO:0043531	ADP binding	MF	8.55E-08	100
	GO:0005199	Structural constituent of cell wall	MF	0.00025097	15
	GO:0071669	Plant-type cell wall organization or biogenesis	BP	0.0061632	18
	GO:0009664	Plant-type cell wall organization	BP	0.013663	17
	GO:0045229	External encapsulating structure organization	BP	0.031101	30
	GO:0071555	Cell wall organization	BP	0.041254	28
	GO:0030312	External encapsulating structure	CC	0.041254	49
DAF60	GO:0004315	3-oxoacyl-[acyl-carrier-protein] synthase activity	MF	0.0066424	5
	GO:0004312	Fatty acid synthase activity	MF	0.0066424	5
	GO:0016747	Transferase activity, transferring acyl groups other than amino-acyl groups	MF	0.012673	12
	GO:0016706	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	MF	0.012673	8
	GO:0016491	Oxidoreductase activity	MF	0.012673	40
	GO:0016705	Oxidoreductase activity, acting on paired donors, with incorporation, or reduction of molecular oxygen	MF	0.016231	15
	GO:0043115	Precorrin-2 dehydrogenase activity	MF	0.016392	5
	GO:0003824	Catalytic activity	MF	0.020356	127
	GO:0008171	O-methyltransferase activity	MF	0.020356	6
	GO:0016628	Oxidoreductase activity, acting on the CH-CH group of donors, NAD or NADP as acceptor	MF	0.020356	6
	GO:0006633	Fatty acid biosynthetic process	BP	0.012673	10
	GO:0072330	Monocarboxylic acid biosynthetic process	BP	0.016392	10
	GO:0019354	Siroheme biosynthetic process	BP	0.016392	5
	GO:0046156	Siroheme metabolic process	BP	0.016392	5
	GO:0008610	Lipid biosynthetic process	BP	0.017694	19
	GO:0006631	Fatty acid metabolic process	BP	0.020356	10
	GO:0044710	Single-organism metabolic process	BP	0.020356	73
	GO:0055114	Oxidation-reduction process	BP	0.028844	38
	GO:0006629	Lipid metabolic process	BP	0.030943	26
	GO:0005835	Fatty acid synthase complex	CC	0.0066424	5

The genes were categorized based on GO.*MF, molecular function; BP, biological process; CC, cellular component.

The *pscb* Mutant Seed Coat Accumulates Phytomelanin Through Higher Level of Polyphenol Oxidases and Peroxidase Expression

Various methods were used to solubilize and characterize the molecule(s) imparting the seed coat color that contribute to the brown pigmentation in *pscb* mutant seed. Compounds comprised of anthocyanins and proanthocyanidin were eliminated as candidates since their contents in *pscb* mutant were much lower than WT. Both bleach and peroxide were capable of removing the brown testa color of the *pscb* mutant seed coat. The intransigent nature of the dark compound, particularly its stability under acid hydrolysis and its susceptibility to the two treatments mentioned above, were reported as hallmarks of melanin, a class of chemically resistant phenolic polymers.

WT and *pscb* seed coat, when hydrolyzed with NaOH, produced little or no precipitate when the hydrolysates were

subsequently acidified to pH 2. Contrasted with hydrolysates, both the *pscb* and WT seed coat produced abundant precipitates upon acidification. Furthermore, the precipitates could be resolubilized in NaOH or in dimethyl sulfoxide (DMSO), consistent with the hypothesis that the black pigment was melanic in nature, means both the WT and *pscb* seed coat had melanin constituents. The melanin content in *pscb* was 64.3 mg/g FW as compared to WT i.e., 27.6 mg/g FW (**Figure 6A**). Among the DEGs of three different stages of *pscb* and WT, there were nine polyphenol oxidases (PPOD) (**Figure 6B**, **Table S4**) and 24 peroxidases (POD) (**Figure 6C**, **Figure S2**, **Table S4**) obviously increased during the late developmental stages especially at DAF60 when compared with the WT. Most of the PPOD and POD showed different expression patterns among the developing process between *pscb* mutant and WT. In the *pscb* mutant, these genes had higher expression level during the seed coat development, while these genes either declined or remained stable in WT.

TABLE 2 | KEGG pathways in common of *pscb* mutant compared with WT at DAF20, DAF40, and DAF60.

KEGG pathway	Gene number		
	DAF20	DAF40	DAF60
Fatty acid elongation	1	6	7
Isoquinoline alkaloid biosynthesis	1	2	1
Tropane, piperidine and pyridine alkaloid biosynthesis	1	2	1
Stilbenoid, diarylheptanoid and gingerol biosynthesis	1	5	1
Tyrosine metabolism	1	8	1
Beta-Alanine metabolism	1	10	1
Cyanoamino acid metabolism	1	10	1
Tryptophan metabolism	1	11	1
Flavonoid biosynthesis	3	11	9
Amino sugar and nucleotide sugar metabolism	3	13	1
Phenylalanine metabolism	2	20	6
Pyrimidine metabolism	1	31	1
Phenylpropanoid biosynthesis	2	33	6
Oxidative phosphorylation	1	38	1
Protein processing in endoplasmic reticulum	1	50	2
Plant hormone signal transduction	1	56	7

Bold means important KEGGs related to the phenotype.

Transcriptional Regulation of ABA and Ethylene Signal Transduction Related Genes during Peanut Seed Coat Development

The KEGG analysis of the RNA-seq data indicated significant change in the expression of ABA and ethylene signal transduction related genes in peanut seed coat, especially the genes of ABA signal transduction pathways. Previous study showed that ABA and ethylene in the maturation process play important roles, and between them there was a very close interaction. To better understand the transcriptional regulation of ABA and ethylene-related genes in peanut seed coat development, genes related to ABA and ethylene signal transduction were analyzed in the three different developing stages of WT and *pscb* mutant seed coat. At the early developing stage (DAF20), there was no DEG related to ABA and ethylene signal transduction. However, at DAF40, there were 13 DEGs in the ABA signal transduction between the *pscb* mutant and WT, including seven PYR/PYL family abscisic acid receptors (down-regulated), five PP2C (three down-regulated and two up-regulated), and two SnRK2 proteins (one up-regulated and one down regulated) (Table S5), showing that the ABA signal transduction was weakened in the mutant. The analysis of ethylene transduction related genes revealed a differential regulation between *pscb* mutant and wild type. The overall expression patterns of the genes in ethylene signal transduction were almost the same which were increased or kept invariant in WT and down-regulated in *pscb* mutant (Table S5). Such as ethylene-responsive transcription factor 1 (*c31399_g2*), the expression level stabilized around 2 (2.68, 1.96, 2.51) FPKM in WT of three stages, while in *pscb* mutant, declined from 1.32 to 0.12 FPKM.

Candidate Genes Related to Peanut Crack and Brown Seed Coat in *pscb*

In order to identify candidate genes controlling seed-coat cracking and seed color, we analyzed common DEGs between WT and *pscb* mutant from the three different developmental stages. The above analysis resulted in identification of 62 unigenes in WT and *pscb* mutant (Figure S1, Table S6). Among the common DEGs, we found three putative candidate genes (*c36498_g1:CCoAOMT1*, *c40902_g2:kinesin*, and *c33560_g1:MYB3*), which were significantly down-regulated in the *pscb* (Table S6). In the seed coat of wild type, the FPKM value of *c36498_g1* gene decreased from 10.96 (DAF20) to 3.38 (DAF40) and then to 8.59. In case of the *pscb* mutant, the FPKM value of *c36498_g1* gene declined from 0.58 (DAF20) to 0.24 (DAF40) and then to 0.03 (DAF60). The expression of *c40902_g2* gene in WT went up gradually from 5.31 (DAF20) to 7.29 (DAF40) and then 18.04 (DAF60). In contrast, the expression level of *c40902_g2* in *pscb* mutant went down from 0.38 (DAF20) to 0.34 (DAF40) and then to 0.17 (DAF60). Very few reads of *c33560_g1* gene were detected in *pscb* mutant, the FPKM were 0.08 (DAF20), 0.07 (DAF40), and 0 (DAF60), while in case of wild type, the FPKM value changed from 5.00 (DAF20) to 2.67 (DAF40) and then 2.77 (DAF60). *C36498_g1* was a caffeoyl-CoA O-methyltransferase which was reported in Phenylalanine metabolism. The *CCoAOMT1* gene from maize, medicago, jute, poplar, *Zinnia elegans*, *Arabidopsis*, and loblolly pine were involved in lignin biosynthesis and cell wall organization (Ye et al., 1994, 2001; Ye and Varner, 1995; Goujon et al., 2003; Zhou et al., 2010; Wagner et al., 2011; Li et al., 2013; Zhang et al., 2014). *C40902_g2* encode a kinesin-4-like protein, and kinesin protein were reported functioned in cell wall organization (Zhong et al., 2002; Zhang et al., 2010; Fujikura et al., 2014; Kong et al., 2015). *c33560_g1* was an R2R3-Myb transcription factor encoding gene. Previous studies showed R2R3-Myb factor worked combined with other transcription factors together to regulate flavonoids and phenylalanine metabolism further regulate the proanthocyanidin biosynthesis (Baudry et al., 2006; Quattrocchio et al., 2006). These genes that might lead to the brown seed color and crack seed coat phenotype need to be confirmed in further functional genomics studies.

DISCUSSION

The conventional understanding of the role of the seed coat is that it provides a protective layer for the developing zygote. It also acts as channel for transmitting environmental cues to the interior of the seed which helps seed to adjust its metabolism in response to changes in its external environment (Radchuk and Borisjuk, 2014). In peanut, flavonoid, and phenylpropanoid biosynthesis pathways were reported to be related with aflatoxin resistance (Garcia et al., 2013; Wang et al., 2016). Therefore, the research on seed coat cracking and pigmentation/color will not only help in understanding and improving the peanut seed quality, it will also help in understanding the genetic control for few seed borne diseases such as aflatoxin contamination. In the present study, RNA-seq was used to investigate the differences

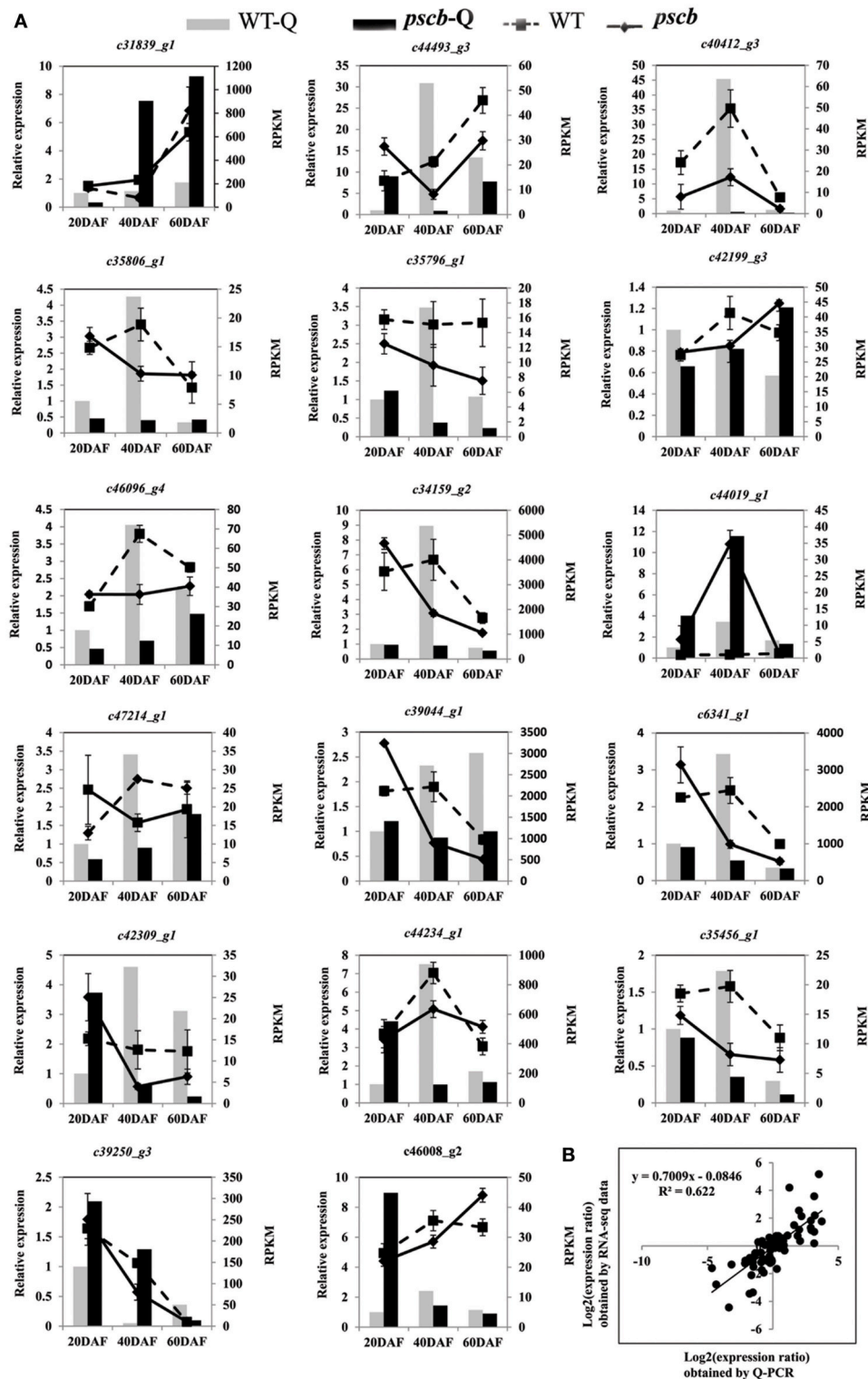
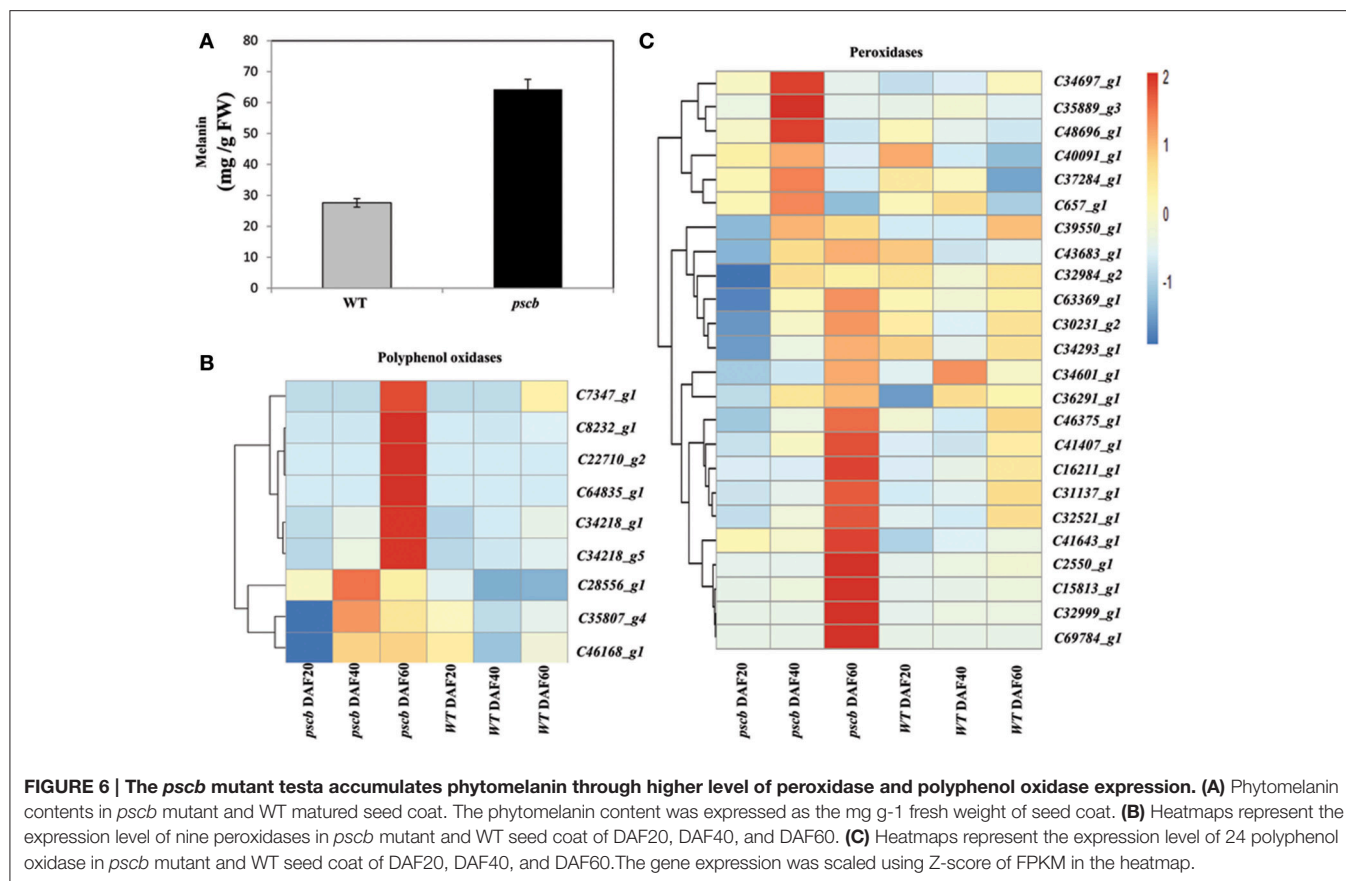


FIGURE 5 | qRT-PCR validation of differential expression. (A) Transcript levels of 17 genes, which were involved in plant cell wall organization **(B)** Comparison between the gene expression ratios obtained from RNA-seq data and qRT-PCR. The RNA-seq log₂ value of the expression ratio (y-axis) has been plotted against the developmental stages (x-axis).



in the transcriptome between the three different stages of *pscb* mutant and its WT. Thousands of genes that were differently regulated during the three stages of seed coat development were identified by transcriptomic profiling.

Seed Coat Pigmentation was Redirected in *pscb* Mutant between Anthocyanin, Proanthocyanidin, Melanin, and Lignin

Flavonoids are the secondary metabolites that accumulate in plants and promote seed and pollen dispersal by contributing to color formation in fruits and flowers (Winkel-Shirley, 2001). Previously, researchers showed that epicatechin derivatives (Marinova et al., 2007; Zhao and Dixon, 2009) and a PA monomer (Holton and Cornish, 1995; Grotewold, 2006) are important flavonols in the synthesis of the seed coat of *Arabidopsis* and *Medicago*. Originally when we first observed the *pscb* mutant, we thought the pigmentation must be markedly enhanced when compared with WT for the sight of deeper color. However, compounds comprised of anthocyanins and lignins were eliminated as candidates, in contrast the production of anthocyanins and proanthocyanidins was reduced in the *pscb* mutant seed coats. We noticed that both bleach and peroxide were capable of removing the dark testa color of the *pscb* mutant seed and the dark compound was stable under acid hydrolysis, these features were reported as hallmarks of melanin (Fogarty and Tobin, 1996; Sava et al., 2001), a class of chemically resistant phenolic polymers. Melanin, was reported as black

pigments, especially in the seed coat of composite, morning glory and many oilseed *Brassica* species (Park et al., 2007; Park and Hoshino, 2012; Yu, 2013). Surprisingly, the melanin content in the *pscb* mutant was more than twice as compared to WT. The seed coat crack and brown color might be contributing for the inhibition of flavonoid and lignin metabolism pathway, and the accumulation of the upper component of the aromatic amino acid converted to be melanin and compensating for the disadvantages of the lower anthocyanin and proanthocyanidin content.

Potential Mechanisms Underlying Seed Coat Color and Crack in Peanut

Flavonoids are plant secondary metabolites derived from the phenylpropanoid pathway. The flavonoid pathway in *Arabidopsis* has been characterized mainly using the mutants. Twenty-three genes have already been identified at the molecular level corresponding to several enzymes (CHS, CHI, F3H, F3'H, DFR, LDOX, FLS, ANR, LACCASE), transports (TT12, TT19, AHA10), and regulatory factors (TT1, TT2, TT8, TT16, TTG1, TTG2, PAP1, GL3, ANL2, FUSCA3, KAN4) (Baxter et al., 2005; Li et al., 2010). The transcriptome data generated in this study showed that the *F3H*, *F3'H*, *DFR*, and *ANR* were suppressed in the stage of DAF40. These results suggested that DAF 40 was the key growth stage for the anthocyanins and proanthocyanidins accumulation. Further, *F3H*, *F3'H*, *DFR*, and *ANR* are the key

genes underlying differences in seed coat pigmentation in *pscb* mutant and WT.

Melanin production resulting in black pigmentation proceeds by one of two pathways in plants. The first leads to compounds of the allomelanin variety deviating in the shikimic acid pathway at β -coumarate before the flavonoids (Goodwin and Mercer, 1983). The second produces eumelanin from Tyr via oxidation of DL-dioxy-Phe and is divorced from the shikimate pathway altogether. Polymerization of melanic compounds produces peroxide (Blois, 1978) and trapped free radicals that should lead to an up-regulation of PRX activity and an EPR signal, respectively. The transcriptome analysis revealed that many structural genes which are involved in the phenylalanine metabolism were found down-regulated in the *pscb* mutant (Figure 7) in addition to decrease of the lignin, anthocyanin, and proanthocyanidin content. In contrast, there were also nine polyphenol oxidases and 24 peroxidases obviously increased at the late developmental stages especially at DAF60 when compared with the WT.

It has already been reported that the down regulation of C3H, 4CL1, F5H, or COMT enzymes in *A. thaliana* affects the final lignin composition (Chapple et al., 1994; Lee et al., 1997; Meyer et al., 1998; Franke et al., 2002; Goujon et al., 2003). The repression of this set of genes in transgenic plant leads to a 70% reduction in the total lignin content and resulted in severe

phenotypic effects. We recently described that the lignin content in the *pscb* mutant decreased by 60%, this strong reduction in lignin content can be related with the down regulation of genes involved in lignin synthesis pathway such as *CAD*, *F5H*, or *COMT*. The above results indicate that the synthesis mechanism of lignin in peanut is similar to other plants.

Previous study hold the idea that lignin is related to PA due to common steps in the phenylpropanoid pathway. The low lignin is found strongly associated with the unpigmented seed coat trait as lignin is usually accumulated within the cell wall whereas PA is usually accumulated in endoplasmic reticulum vesicles or in the plant vacuole. Lignin variability may influence seed coat pigment extractability, owing to the position of the highly lignified palisade cells adjacent to the inner integument in the seed coat, where pigment is initially deposited (Marles and Gruber, 2004). Here, our result showed that the anthocyanins, proanthocyanidins, and the lignin content were reduced while the melanin content was enhanced, indicating new regulation mechanism may exist in peanut seed coat.

Regulation Mechanisms Underlying the Biosynthesis of Seed Coat Pigment in Peanut

Plant hormone and transcription factors play an important role in plant seed development. Through the analysis of the

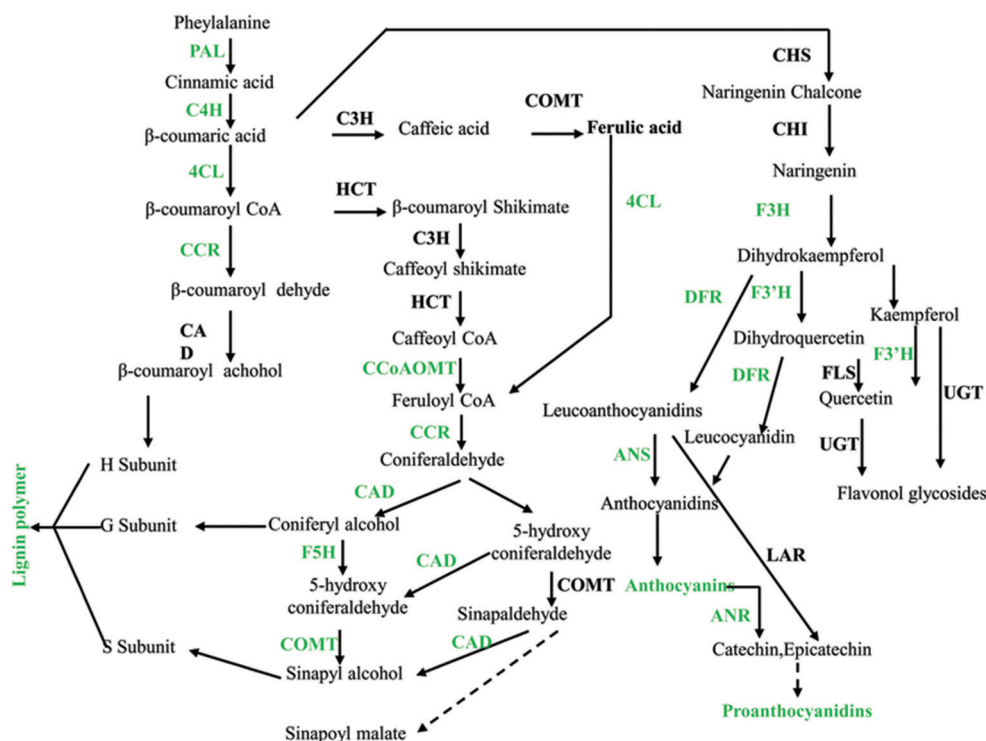
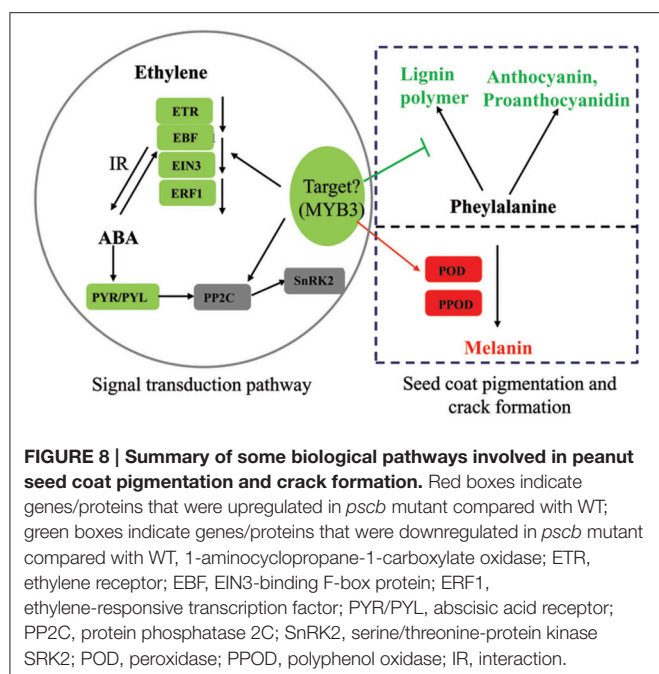


FIGURE 7 | Phenylalanine metabolism was down-regulated in *pscb* mutant seed coat. PAL, phenylalanine ammonia-lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate-CoA ligase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; HCT, hydroxycinnamoyl-CoA shikimate/quinic acid hydroxycinnamoyl transferase; C3H, 4-coumarate 3-hydroxylase; COMT, caffeic acid o-methyltransferase; CCoAOMT, caffeoyl-CoA o-methyltransferase; F5H, ferulate-5-hydroxylase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F3'H, flavonoid 3'-hydroxylase; FLS, flavonol synthase; UGTs, UDP sugar glycosyltransferases; DFR, dihydroflavonol reductase.



RNA-seq data, we found that hormone-related genes in the *pscb* mutant seed coat during seed development changed greatly when compared with WT especially the ABA and ethylene signal transduction pathway. Plant hormone ABA has been suggested to play a role in fruit anthocyanin biosynthesis (Shen et al., 2014; Li et al., 2015). Our transcriptome data showed that all the seven ABA receptor PYR/PYL genes, which presented in the KEGG “plant hormone signal transduction,” were obviously down regulated in the *pscb* mutant indicating the ABA signal transduction might be strongly suppressed in the mutant leading to the lower anthocyanin and proanthocyanidin content. Ethylene is required for the onset of accumulation of anthocyanins (Chervin et al., 2004). In this study, one ETR1, three EBF1, two EIN3, and two ERF1 involved in ethylene signal transduction were down regulated at DAF40 in *pscb* mutant. Further, the three EBF1 and two ERF1 were also down-regulated in the stage of DAF60, which strongly demonstrate that the ethylene signaling was suppressed in the mutant. MYB transcription factors have been well-reported in the regulation of plant pigmentation in different species (Appelhaugen et al., 2011; Liu et al., 2014; An et al., 2015; Cavallini et al., 2015; Yoshida et al., 2015). Through analysis of the common DEGs of the three different stages, we identified a R2R3-MYB transcription factor (*c33560_g1*), with extremely low expression in the stages of DAF20 and DAF40 and none expression in the late stage of DAF60 in *pscb* mutant. By blasting the protein sequence of *c33560_g1* in The *Arabidopsis* Information Resource (TAIR), AT1G22640 (MYB3) showed the highest similarity, while AT1G22640 was reported as an MYB-type transcription factor (MYB3) that represses phenylpropanoid biosynthesis gene expression (Rowan et al., 2009). The MdMYB3, the homolog in apple, was similarly identified as regulator of anthocyanin biosynthesis and

flower development (Vimolmangkang et al., 2013), indicating the potential function of *c33560_g1* in peanut seed coat pigmentation.

In the present study, the ABA pathway, ethylene pathway, and the R2R3-MYB transcription factor (*c33560_g1*) were all different between *pscb* mutant and WT and were selected as important candidate for the format of the cracking and brown seed coat phenotype. We hypothesize that the R2R3-MYB transcription factor (*c33560_g1*), ABA and ethylene signaling pathways interact cooperatively to suppress the anthocyanin, proanthocyanidin, and lignin synthesis related pathways' genes to influence the anthocyanin, proanthocyanidin and lignin level. Simultaneously, the enhanced expression of POD and PPOD encoding genes further regulate seed coat pigmentation (Figure 8). The above does not prove the interaction relationship of the three factors responsible for seed-coat cracking and brown seed color trait of *pscb* mutant and requires further detailed study.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: LW, YL, LY, HJ, BSL. Performed the experiments: LW, YW, BL. Analyzed the data: LW, MP. Contributed reagents/materials/analysis tools: GW, XD, YL, LY, HJ. Wrote the paper: LW, MP, RV, BSL. All authors have read and approved the manuscript.

ACKNOWLEDGMENTS

This work was funded by the National Basic Research Program of China (2013CB127803), the National Natural Science Foundation of China (No. 31461143022, 31371662, 31301256), and the China Agriculture Research System (No. CARS-14). This work has been also undertaken as part of the CGIAR Research Program on Grain Legumes and the China-CGIAR research collaborative project. ICRISAT is a member of CGIAR Consortium.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01491>

Figure S1 | Venn diagram showing the number of genes with increased and decreased expression of three different development stages.

Figure S2 | qPCR verification of the POD AND PPOD change in *pscb* mutant compared with WT.

Table S1 | DEGs between *pscb* and WT at DAF20, DAF40, and DAF60.

Table S2 | KEGG pathway of the DEGs between *pscb* mutant and WT at DAF20, DAF40, and DAF60.

Table S3 | Gene-specific primers for real time PCR analysis.

Table S4 | Expression and annotation the differential expressed POD and PPOD genes.

Table S5 | Differentially expressed genes involved in ABA and ethylene signal transduction between *pscb* mutant and WT.

Table S6 | Common DEGs between WT and *pscb* mutant of the three different developmental stages.

REFERENCES

- Abrahams, S., Lee, E., Walker, A. R., Tanner, G. J., Larkin, P. J., and Ashton, A. R. (2003). The *Arabidopsis* *TDS4* gene encodes leucoanthocyanidin dioxygenase (LDOX) and is essential for proanthocyanidin synthesis and vacuole development. *Plant J.* 35, 624–636. doi: 10.1046/j.1365-313X.2003.01834.x
- Agarwal, K., and Menon, K. (1974). Lignin content and seed coat thickness in relation to seed coat cracking in soybean. *Seed Res.* 2, 64–66.
- An, X. H., Tian, Y., Chen, K. Q., Liu, X. J., Liu, D. D., Xie, X. B., et al. (2015). MdMYB9 and MdMYB11 are involved in the regulation of the JA-induced biosynthesis of anthocyanin and proanthocyanidin in apples. *Plant Cell Physiol.* 56, 650–662. doi: 10.1093/pcp/pcu205
- Anterola, A. M., and Lewis, N. G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. *Phytochemistry* 61, 221–294. doi: 10.1016/S0031-9422(02)00211-X
- Appelhaagen, I., Lu, G. H., Huep, G., Schmelzer, E., Weisshaar, B., and Sagasser, M. (2011). TRANSPARENT TESTA1 interacts with R2R3-MYB factors and affects early and late steps of flavonoid biosynthesis in the endothelium of *Arabidopsis thaliana* seeds. *Plant J.* 67, 406–419. doi: 10.1111/j.1365-313X.2011.04603.x
- Ballard, T. S., Mallikarjunan, P., Zhou, K., and O'Keefe, S. F. (2009). Optimizing the extraction of phenolic antioxidants from peanut skins using response surface methodology. *J. Agric. Food Chem.* 57, 3064–3072. doi: 10.1021/jf8030925
- Baudry, A., Caboche, M., and Lepiniec, L. (2006). TT8 controls its own expression in a feedback regulation involving TTG1 and homologous MYB and bHLH factors, allowing a strong and cell-specific accumulation of flavonoids in *Arabidopsis thaliana*. *Plant J.* 46, 768–779. doi: 10.1111/j.1365-313X.2006.02733.x
- Baxter, I. R., Young, J. C., Armstrong, G., Foster, N., Bogenschütz, N., Cordova, T., et al. (2005). A plasma membrane H⁺-ATPase is required for the formation of proanthocyanidins in the seed coat endothelium of *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 2649–2654. doi: 10.1073/pnas.0406377102
- Blois, M. S. (1978). The melanins: their synthesis and structure. *Photochem. Photobiol.* 3, 115–134. doi: 10.1007/978-1-4684-2580-2_3
- Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54, 519–546. doi: 10.1146/annurev.arplant.54.031902.134938
- Cavallini, E., Matus, J. T., Finezzo, L., Zenoni, S., Loyola, R., Guzzo, F., et al. (2015). The phenylpropanoid pathway is controlled at different branches by a set of R2R3-MYB C2 repressors in grapevine. *Plant Physiol.* 167, 1448–1470. doi: 10.1104/pp.114.256172
- Chapple, C. C. S., Shirley, B. W., Zook, M., Hammerschmidt, R., and Somerville, S. C. (1994). "Secondary metabolism in *Arabidopsis*," in *Arabidopsis*, eds E. M. Meyerowitz, and C. R. Somerville (Plainview: Cold Spring Harbor Laboratory Press), 989–1030.
- Chervin, C., El-Kereamy, A., Roustan, J. P., Latche, A., Lamon, J., and Bouzayen, M. (2004). Ethylene seems required for the berry development and ripening in grape, a non-climacteric fruit. *Plant Sci.* 167, 1301–1305. doi: 10.1016/j.plantsci.2004.06.026
- Debeaujon, I., Léon-Kloosterziel, K. M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol.* 122, 403–414. doi: 10.1104/pp.122.2.403
- de Camargo, A. C., Vidal, C. M., Canniatti-Brazaca, S. G., and Shahidi, F. (2014). Fortification of cookies with peanut skins: effects on the composition, polyphenols, antioxidant properties, and sensory quality. *J. Agric. Food Chem.* 62, 11228–11235. doi: 10.1021/jf503625p
- Devic, M., Guillemot, J., Debeaujon, I., Bechtold, N., Bensaude, E., Koornneef, M., et al. (1999). The BANYULS gene encodes a DFR-like protein and is a marker of early seed coat development. *Plant J.* 19, 387–398. doi: 10.1046/j.1365-313X.1999.00529.x
- Dixon, R. A., Xie, D. Y., and Sharma, S. B. (2005). Proanthocyanidins—a final frontier in flavonoid research? *New Phytol.* 165, 9–28. doi: 10.1111/j.1469-8137.2004.01217.x
- Dubos, C., Le Gourrier, J., Baudry, A., Huep, G., Lanet, E., Debeaujon, I., et al. (2008). MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant J.* 55, 940–953. doi: 10.1111/j.1365-313X.2008.03564.x
- Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C., and Lepiniec, L. (2010). MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* 15, 573–581. doi: 10.1016/j.tplants.2010.06.005
- Duke, S. H., Kakefuda, G., and Harvey, T. M. (1983). Differential leakage of intracellular substances from imbibing seeds. *Plant Physiol.* 72, 919–924. doi: 10.1104/pp.72.4.919
- Duke, S. H., Kakefuda, G., Henson, C. A., Loeffler, N. L., and van Hulle, N. M. (1986). Role of the testa epidermis in the leakage of intracellular substances from imbibing soybean seeds and its implications for seedling survival. *Physiol. Plant.* 68, 625–631. doi: 10.1111/j.1399-3054.1986.tb03407.x
- Fogarty, R. V., and Tobin, J. M. (1996). Fungal melanins and their interactions with metals. *Enzyme Microb. Technol.* 19, 311–317.
- Fornalé, S., Shi, X., Chai, C., Encina, A., Irar, S., Capellades, M., et al. (2010). ZmMYB31 directly represses maize lignin genes and redirects the phenylpropanoid metabolic flux. *Plant J.* 64, 633–644. doi: 10.1111/j.1365-313X.2010.04363.x
- Franke, R., Hemm, M. R., Denault, J. W., Ruegger, M. O., Humphreys, J. M., and Chapple, C. (2002). Changes in secondary metabolism and deposition of an unusual lignin in the *ref8* mutant of *Arabidopsis*. *Plant J.* 30, 47–59. doi: 10.1046/j.1365-313X.2002.01267.x
- Fujikura, U., Elsaesser, L., Breuninger, H., Sánchez-Rodríguez, C., Ivakov, A., Laux, T., et al. (2014). Atkinesin-13A modulates cell-wall synthesis and cell expansion in *Arabidopsis thaliana* via the THESEUS1 pathway. *PLoS Genet.* 10:e1004627. doi: 10.1371/journal.pgen.1004627
- García, D., Ramos, A. J., Sanchis, V., and Marin, S. (2013). Equisetum arvense hydro-alcoholic extract: phenolic composition and antifungal and antimycotoxigenic effect against *Aspergillus flavus* and *Fusarium verticillioides* in stored maize. *J. Sci. Food Agric.* 93, 2248–2253. doi: 10.1002/jsfa.6033
- Goodwin, T. W., and Mercer, E. I. (1983). *Introduction to Plant Biochemistry*, Edn. 2. Oxford: Pergamon Press.
- Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., Lu, F., et al. (2003). A new *Arabidopsis thaliana* mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. *Plant Mol. Biol.* 51, 973–989. doi: 10.1023/A:1023022825098
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Grotewold, E. (2006). The genetics and biochemistry of floral pigments. *Annu. Rev. Plant Biol.* 57, 761–780. doi: 10.1146/annurev.arplant.57.032905.105248
- Hafez, A. A. A., Gaafer, A. K., and Allam, A. M. M. (1981). Inheritance of flesh colour, seed coat cracks and total soluble solids in watermelon and their genetic relations. I. Quantitative traits. *Acta Agron. Acad. Sci.* 30, 82–86.
- Hoebler, C., Barry, J.-L., David, A., and Delort-Laval, J. (1989). Rapid acid hydrolysis of plant cell wall polysaccharides and simplified quantitative determination of their neutral monosaccharides by gas-liquid chromatography. *J. Agric. Food Chem.* 37, 360–365. doi: 10.1021/jf00086a020
- Holton, T. A., and Cornish, E. C. (1995). Genetics and biochemistry of anthocyanin biosynthesis. *Plant Cell* 7, 1071–1083. doi: 10.1105/tpc.7.7.1071
- Kirk, T. K., and Obst, J. R. (1988). Lignin determination. *Method Enzymol.* 161, 87–101. doi: 10.1016/0076-6879(88)61014-7
- Kong, Z., Ioki, M., Braybrook, S., Li, S., Ye, Z. H., Julie Lee, Y. R., et al. (2015). Kinesin-4 functions in vesicular transport on cortical microtubules and regulates cell wall mechanics during cell elongation in Plants. *Mol. Plant* 8, 1011–1023. doi: 10.1016/j.molp.2015.01.004
- Kour, A., Boone, A. M., and Vodkin, L. O. (2014). RNA-Seq profiling of a defective seed coat mutation in Glycine max reveals differential expression of proline-rich and other cell wall protein transcripts. *PLoS ONE* 9:e96342. doi: 10.1371/journal.pone.0096342
- Lee, D., Meyer, K., Chapple, C., and Douglas, C. J. (1997) Antisense suppression of 4-coumarate:coenzyme a ligase activity in *Arabidopsis* leads to altered lignin subunit composition. *Plant Cell* 9, 1985–1998. doi: 10.1105/tpc.9.11.1985
- Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Li, Q., Chen, P., Dai, S., Sun, Y., Yuan, B., Kai, W., et al. (2015). PacCYP707A2 negatively regulates cherry fruit ripening while PacCYP707A1 mediates drought tolerance. *J. Exp. Bot.* 66, 3765–3774. doi: 10.1093/jxb/erv169

- Li, X., Chen, W., Zhao, Y., Xiang, Y., Jiang, H., Zhu, S., et al. (2013). Downregulation of caffeoyl-CoA O-methyltransferase (CCoAOMT) by RNA interference leads to reduced lignin production in maize straw. *Genet. Mol. Biol.* 36, 540–546. doi: 10.1590/S1415-47572013005000039
- Li, X., Westcott, N., Links, M., and Gruber, M. Y. (2010). Seed coat phenolics and the developing silique transcriptome of *Brassica carinata*. *J. Agric. Food Chem.* 58, 10918–10928. doi: 10.1021/jf102208a
- Liu, C., Jun, J. H., and Dixon, R. A. (2014). MYB5 and MYB14 Play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiol.* 165, 1424–1439. doi: 10.1104/pp.114.241877
- Liu, H. L. (1949). Inheritance of defective seed coat in soybeans. *J. Hered.* 40, 317–322.
- Ma, Y., Kerr, W. L., Swanson, R. B., Hargrove, J. L., and Pegg, R. B. (2014). Peanut skins-fortified peanut butters: effect of processing on the phenolics content, fibre content and antioxidant activity. *Food Chem.* 145, 883–891. doi: 10.1016/j.foodchem.2013.08.125
- Mao, X., Cai, T., Olyarchuk, J. G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430
- Marinova, K., Kleinschmidt, K., Weissenböck, G., and Klein, M. (2007). Flavonoid biosynthesis in barley primary leaves requires the presence of the vacuole and controls the activity of vacuolar flavonoid transport. *Plant Physiol.* 144, 432–444. doi: 10.1104/pp.106.094748
- Marles, M. A. S., and Gruber, M. Y. (2004). Histochemical characterisation of unextractable seed coat pigments and quantification of extractable lignin in the *Brassicaceae*. *J. Sci. Food Agric.* 84, 251–262. doi: 10.1002/jsfa.1621
- Mellerowicz, E. J., Baucher, M., Sundberg, B., and Boerjan, W. (2001). Unravelling cell wall formation in the woody dicot stem. *Plant Mol. Biol.* 47, 239–274. doi: 10.1023/A:1010699919325
- Meyer, K., Shirley, A. M., Cusumano, J. C., Bell-Lelong, D. A., and Chapple, C. (1998). Lignin monomer composition is determined by the expression of a cytochrome P450-dependent monooxygenase in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 95, 6619–6623. doi: 10.1073/pnas.95.12.6619
- Pang, Y., Wenger, J. P., Saathoff, K., Peel, G. J., Wen, J., Huhman, D., et al. (2009). A WD40 repeat protein from *Medicago truncatula* is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. *Plant Physiol.* 151, 1114–1129. doi: 10.1104/pp.109.144022
- Park, K. I., and Hoshino, A. (2012). A WD40-repeat protein controls proanthocyanidin and phytomelanin pigmentation in the seed coats of the Japanese morning glory. *Plant Physiol.* 169, 523–528. doi: 10.1016/j.jplph.2011.12.001
- Park, K. I., Ishikawa, N., Morita, Y., Choi, J. D., Hoshino, A., and Iida, S. (2007). A bHLH regulatory gene in the common morning glory, *Ipomoea purpurea*, controls anthocyanin biosynthesis in flowers, proanthocyanidin and phytomelanin pigmentation in seeds, and seed trichome formation. *Plant J.* 49, 641–654. doi: 10.1111/j.1365-313X.2006.02988.x
- Quattrocchio, F., Verweij, W., Kroon, A., Spelt, C., Mol, J., and Koes, R. (2006). PH4 of *Petunia* is an R2R3 MYB protein that activates vacuolar acidification through interactions with basic-helix-loop-helix transcription factors of the anthocyanin pathway. *Plant Cell* 18, 1274–1291. doi: 10.1105/tpc.105.034041
- Radchuk, V., and Borisjuk, L. (2014). Physical, metabolic and developmental functions of the seed coat. *Front. Plant Sci.* 5:510. doi: 10.3389/fpls.2014.00510
- Rowan, D. D., Cao, M., Lin-Wang, K., Cooney, J. M., Jensen, D. J., Austin, P. T., et al. (2009). Environmental regulation of leaf colour in red 35S:PAP1 *Arabidopsis thaliana*. *New Phytol.* 182, 102–115. doi: 10.1111/j.1469-8137.2008.02737.x
- Sava, V. M., Yang, S.-M., Hong, M.-Y., Yang, P.-C., and Huang, G. S. (2001). Isolation and characterization of melanic pigments derived from tea and tea polyphenols. *Food Chem.* 73, 177–184. doi: 10.1016/S0308-8146(00)00258-2
- Schlub, R. L., and Schmitthenner, A. F. (1978). Effects of soybean seed coat crackson seed exudation and seedling quality in soil infested with *Pythium ultimum*. *Phytopathology* 68, 1186–1191. doi: 10.1094/Phyto-68-1186
- Shen, X., Zhao, K., Liu, L., Zhang, K., Yuan, H., Liao, X., et al. (2014). A role for PacMYBA in ABA-regulated anthocyanin biosynthesis in red-colored sweet cherry cv. Hong Deng (*Prunus avium* L.). *Plant Cell Physiol.* 55, 862–880. doi: 10.1093/pcp/pcu013
- Shirley, B. W., Kubasek, W. L., Storz, G., Bruggemann, E., Koornneef, M., Ausubel, F. M., et al. (1995). Analysis of *Arabidopsis* mutants deficient in flavonoid biosynthesis. *Plant J.* 8, 659–671. doi: 10.1046/j.1365-313X.1995.08050659.x
- Stewart, R. T., and Wentz, J. B. (1930). A defective seed-coat character in soybeans. *J. Am. Soc. Agron.* 22, 658–662. doi: 10.2134/agronj1930.00021962002200070011x
- Stracke, R., Favory, J. J., Gruber, H., Bartelniewoehner, L., Bartels, S., Binkert, M., et al. (2010a). The *Arabidopsis* bZIP transcription factor HY5 regulates expression of the *PFG1/MYB12* gene in response to light and ultraviolet-B radiation. *Plant Cell Environ.* 33, 88–103. doi: 10.1111/j.1365-3040.2009.02061.x
- Stracke, R., Jahns, O., Keck, M., Tohge, T., Niehaus, K., Fernie, A. R., et al. (2010b). Analysis of PRODUCTION OF FLAVONOL GLYCOSIDES-dependent flavonol glycoside accumulation in *Arabidopsis thaliana* plants reveals MYB11-, MYB12- and MYB111-independent flavonol glycoside accumulation. *New Phytol.* 188, 985–1000. doi: 10.1111/j.1469-8137.2010.03421.x
- Tanaka, Y., Sasaki, N., and Ohmiya, A. (2008). Biosynthesis of plant pigments: anthocyanins, betalains and carotenoids. *Plant J.* 54, 733–749. doi: 10.1111/j.1365-313X.2008.03447.x
- Vimolmangkang, S., Han, Y., Wei, G., and Korban, S. S. (2013). An apple MYB transcription factor, MdMYB3, is involved in regulation of anthocyanin biosynthesis and flower development. *BMC Plant Biol.* 13:176. doi: 10.1186/1471-2229-13-176
- Wagner, A., Tobimatsu, Y., Phillips, L., Flint, H., Torr, K., Donaldson, L., et al. (2011). CCoAOMT suppression modifies lignin composition in *Pinus radiata*. *Plant J.* 67, 119–129. doi: 10.1111/j.1365-313X.2011.04580.x
- Wang, H., Lei, Y., Wan, L., Yan, L., Lv, J., Dai, X., et al. (2016). Comparative transcript profiling of resistant and susceptible peanut post-harvest seeds in response to aflatoxin production by *Aspergillus flavus*. *BMC Plant Biol.* 16:54. doi: 10.1186/s12870-016-0738-z
- Wang, J., Yuan, X., Jin, Z., Tian, Y., and Song, H. (2007). Free radical and reactive oxygen species scavenging activities of peanut skins extract. *Food Chem.* 104, 242–250. doi: 10.1016/j.foodchem.2006.11.035
- Weber, H., Borisjuk, L., and Wobus, U. (2005). Molecular physiology of legume seed development. *Annu. Rev. Plant Biol.* 56, 253–279. doi: 10.1146/annurev.arplant.56.032604.144201
- Winkel-Shirley, B. (2001). Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiol.* 126, 485–493. doi: 10.1104/pp.126.2.485
- Wolf, W. J., and Baker, F. L. (1972). Scanning electron microscopy of soybeans. *Cereal Sci. Today* 17, 125–130.
- Woodworth, C. M., and Williams, L. F. (1938). Recent studies on the genetics of the soybean. *J. Am. Soc. Agron.* 30, 125–129. doi: 10.2134/agronj1938.00021962003000020006x
- Ye, Z. H., Kneusel, R. E., Matern, U., and Varner, J. E. (1994). An alternative methylation pathway in lignin biosynthesis in *Zinnia*. *Plant Cell* 6, 1427–1439. doi: 10.1105/tpc.6.10.1427
- Ye, Z. H., and Varner, J. E. (1995). Differential expression of two O-methyltransferases in lignin biosynthesis in *Zinnia elegans*. *Plant Physiol.* 108, 459–467.
- Ye, Z. H., Zhong, R., Morrison, W. H., and Himmelsbach, D. S. (2001). Caffeoyl coenzyme A O-methyltransferase and lignin biosynthesis. *Phytochemistry* 57, 1177–1185. doi: 10.1016/S0031-9422(01)00051-6
- Yoshida, K., Ma, D., and Constabel, C. P. (2015). The MYB182 protein down-regulates proanthocyanidin and anthocyanin biosynthesis in poplar by repressing both structural and regulatory flavonoid genes. *Plant Physiol.* 167, 693–710. doi: 10.1104/pp.114.253674
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14
- Yu, C. Y. (2013). Molecular mechanism of manipulating seed coat coloration in oilseed Brassica species. *J. Appl. Genet.* 54, 135–145. doi: 10.1007/s13353-012-0132-y
- Zhang, G., Zhang, Y., Xu, J., Niu, X., Qi, J., Tao, A., et al. (2014). The CCoAOMT1 gene from jute (*Corchorus capsularis* L.) is involved in lignin biosynthesis in *Arabidopsis thaliana*. *Gene* 546, 398–402. doi: 10.1016/j.gene.2014.05.011

- Zhang, H., Yerigui, Y. Y., and Ma, C. (2013). Structures and antioxidant and intestinal disaccharidase inhibitory activities of A-type proanthocyanidins from peanut skin. *J. Agric. Food Chem.* 61, 8814–8820. doi: 10.1021/jf402518k
- Zhang, M., Zhang, B., Qian, Q., Yu, Y., Li, R., Zhang, J., et al. (2010). Brittle Culm 12, a dual-targeting kinesin-4 protein, controls cell-cycle progression and wall properties in rice. *Plant J.* 63, 312–328. doi: 10.1111/j.1365-313X.2010.04238.x
- Zhao, J., and Dixon, R. A. (2009). MATE transporters facilitate vacuolar uptake of epicatechin 3'-O-glucoside for proanthocyanidin biosynthesis in *Medicago truncatula* and *Arabidopsis*. *Plant Cell* 21, 2323–2340. doi: 10.1105/tpc.109.067819
- Zhong, R., Burk, D. H., Morrison, W. H., and Ye, Z. H. (2002). A kinesin-like protein is essential for oriented deposition of cellulose microfibrils and cell wall strength. *Plant Cell* 14, 3101–3117. doi: 10.1105/tpc.005801
- Zhou, R., Jackson, L., Shadle, G., Nakashima, J., Temple, S., Chen, F., et al. (2010). Distinct cinnamoyl CoA reductases involved in parallel routes to lignin in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17803–17808. doi: 10.1073/pnas.1012900107
- Zhu, H. F., Fitzsimmons, K., Khandelwal, A., and Kranz, R. G. (2009). CPC, a single-repeat R3 MYB, is a negative regulator of anthocyanin biosynthesis in *Arabidopsis*. *Mol. Plant* 2, 790–802. doi: 10.1093/mp/ssp030

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Wan, Li, Pandey, Wu, Lei, Yan, Dai, Jiang, Zhang, Wei, Varshney and Liao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Evaluation of Exotically-Derived Soybean Breeding Lines for Seed Yield, Germination, Damage, and Composition under Dryland Production in the Midsouthern USA

Nacer Bellaloui^{1*}, James R. Smith¹, Alemu Mengistu², Jeffery D. Ray¹ and Anne M. Gillen¹

¹ Crop Genetics Research Unit, USDA Agricultural Research Service, Stoneville, MS, USA, ² Crop Genetics Research Unit, USDA Agricultural Research Service, Jackson, TN, USA

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica—Universidade Nova de
Lisboa, Portugal

Reviewed by:

Elisa Bellucci,
Università Politecnica delle Marche,
Italy
Hao Peng,
Washington State University, USA

*Correspondence:

Nacer Bellaloui
nacer.bellaloui@ars.usda.gov

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 08 September 2016

Accepted: 27 January 2017

Published: 27 February 2017

Citation:

Bellaloui N, Smith JR, Mengistu A,
Ray JD and Gillen AM (2017)
Evaluation of Exotically-Derived
Soybean Breeding Lines for Seed
Yield, Germination, Damage, and
Composition under Dryland
Production in the Midsouthern USA.
Front. Plant Sci. 8:176.
doi: 10.3389/fpls.2017.00176

Although the Early Soybean Production System (ESPS) in the Midsouthern USA increased seed yield under irrigated and non-irrigated conditions, heat stress and drought still lead to poor seed quality in heat sensitive soybean cultivars. Our breeding goal was to identify breeding lines that possess high germination, nutritional quality, and yield potential under high heat and dryland production conditions. Our hypothesis was that breeding lines derived from exotic germplasm might possess physiological and genetic traits allowing for higher seed germinability under high heat conditions. In a 2-year field experiment, breeding lines derived from exotic soybean accessions, previously selected for adaptability to the ESPS in maturity groups (MG) III and IV, were grown under non-irrigated conditions. Results showed that three exotic breeding lines had consistently superior germination across 2 years. These lines had a mean germination percentage of >80%. Two (25-1-1-4-1-1 and 34-3-1-2-4-1) out of the three lines with ≥80% germination in both years maintained high seed protein, oleic acid, N, P, K, B, Cu, and Mo in both years. Significant ($P < 0.05$) positive correlations were found between germination and oleic acid and with K and Cu in both years. Significant negative correlations were found between germination and linoleic acid, Ca, and hard seed in both years. There were positive correlations between germination and N, P, B, Mo, and palmitic acid only in 2013. A negative correlation was found between germination and green seed damage and linolenic acid in 2013 only. Seed wrinkling was significantly negatively correlated with germination in 2012 only. A lower content of Ca in the seed of high germinability genotypes may explain the lower rates of hard seed in those lines, which could lead to higher germination. Many of the differences in yield, germination, diseases, and seed composition between years are likely due to heat and rainfall differences between years. The results also showed the potential roles of seed minerals, especially K, Ca, B, Cu, and Mo, in maintaining high seed quality. The knowledge gained from this research will help breeders to select for soybean with high seed nutritional qualities and high germinability.

Keywords: soybean nutrition, seed composition, mineral nutrition, seed protein, seed oil, germination, seed diseases

INTRODUCTION

The development of the Early Soybean Production System (ESPS) in the Midsouthern resulted in higher yield under irrigated and non-irrigated conditions (Heatherly, 1999). However, high heat and drought in the ESPS are still major environmental stress factors, resulting in poor seed quality (Mengistu and Heatherly, 2006; Smith et al., 2008; Mengistu et al., 2009, 2010) and yield reduction for heat sensitive soybeans, especially under dryland conditions. In addition, early-maturing cultivars grown in the Midsouth USA are often exposed to higher temperature, rainfall, and relative humidity, resulting in high seed infection (TeKrony et al., 1980, 1984) from pathogens such as *Phomopsis* spp. (Kmetz et al., 1974, 1978, 1979), reduction of seed quality (low viability, moldy seed, and reduced emergence; Kmetz et al., 1978; TeKrony et al., 1980), and lower market grade and reduced quality of meal and oil (Hepperly and Sinclair, 1978).

Previous research showed that high temperature and high humidity promote the development of seed with substandard germination and poor seed quality due to diseases such as *Phomopsis longicolla* Hobbs (Thomison et al., 1990; Tekrony et al., 1996; Mengistu and Heatherly, 2006), seed coat wrinkling (Franca-Neto et al., 1988); seed coat shriveling (Franca-Neto et al., 1993; Spears et al., 1997), weathering (Keith and Delouche, 1999), and hard seed (impermeable seed coat; Gibson and Mullen, 1996; Spears et al., 1997; Kebede et al., 2014). Identifying soybean lines with heat-tolerance under dryland conditions could be an effective way to further optimize seed yield and maintain high seed quality (viability, germination, vigor, and composition).

Germinability (germination and vigor) is an important trait for seed producers, and seed composition (seed protein, oil, fatty acids, and mineral nutrition) is important for seed consumers. For example, in Mississippi the minimum germination rate required for certified seed is 80%, and seed lots with less than a 60% germination rate are illegal to sell (Keith and Delouche, 1999). High germination is essential for adequate stand establishment and successful crop production. Previous research reported that the ancestors of modern soybean cultivars in the USA lack high germinability (Smith et al., 2008). Without the introgression of new genetic diversity from exotic germplasm into the breeding gene pool used by commercial seed companies, the new cultivars of the future may also lack high germinability. Smith et al. (2008) identified soybean germplasm accessions with high seed germinability for seed produced under high temperature environments in the ESPS of the Midsouthern USA. They reported that 63 accessions were identified as having a mean standard field germination of $\geq 90\%$ as well as $< 10\%$ hard seediness, *P. longicolla* infection and wrinkled seed coat. They were able to identify genotypes with seed traits that can be used in a breeding program to develop cultivars with high seed germinability for use under high temperature production environments such as in ESPS. Salmeron et al. (2014) studied maturity group choices for early and late planting dates under Midsouthern environments using eight locations in 2012 and 10 locations in 2013, four planting dates and 16 cultivars of maturity MG III through VI. They showed that MG IV and V cultivars

had higher average yield in early-planting systems, but late MG III to late MG IV cultivars had higher yield in late-planting. It was explained that the main characteristic of the better yielders, for example MG IV cultivars, was that the cultivars were more stable across different environments for early and late planting, and that there was a reduced risk for low yield (Salmeron et al., 2014).

Seed composition is another critical quality trait because soybean seed is a major source of protein and oil (saturated fatty acids such as palmitic and stearic, and unsaturated fatty acids such as oleic, linoleic, and linolenic; Wilson, 2004). Also, soybean seed contains macro- and micro-minerals (Sale and Campbell, 1980; Bellaloui et al., 2015). Comparative studies of soybean seed quality among producing countries showed that US soybeans and soybean meal have lower protein contents than Brazil but higher protein than Argentina (Karr-Lilienthal et al., 2004; Thakur and Hurburgh, 2007), affecting the global competitive market of US soybean. On the other hand, US soybeans had the highest concentration of total essential amino acids, making US soybeans superior in protein quality compared to Brazilian and Chinese cultivars (Grieshop and Fahey, 2001; Karr-Lilienthal et al., 2004; Oltmans-Deardorff et al., 2013). A study of 105 soybean genotypes indicated that US cultivars had on average 41.3% protein and 19.9% oil content on the seed dry mass; however, Japanese and South Korean cultivars contained on average 44.5% protein and 18.1% oil (Shi et al., 2010). This difference is due to different genetic backgrounds (Shi et al., 2010), water availability (Rotundo and Westgate, 2009; Rotundo et al., 2014), temperature (Dornbos and Mullen, 1992; Piper and Boote, 1999; Bellaloui et al., 2009a), and soil fertility (Nakasathien et al., 2000; Ray et al., 2006; Bellaloui et al., 2009b). Therefore, improvement of soybean seed composition for protein and oil content has been critical for almost a decade (Durham, 2003). To address soybean seed composition quality, the United Soybean Board initiated the Better Bean Initiative (BBI) and its Technology Utilization Center (TUC) to improve soybean composition and to keep U.S. soybeans competitive in the world market.

One of the goals of the Better Bean Initiative, launched in 2000, was to modify the ratios of fatty acids (high oleic and low linolenic) in oil processing because high oleic and low linolenic acids contribute to the oxidative stability of the oil and improved shelf-life. It was suggested that the most desirable phenotype for soybean oil is $< 7\%$ saturates (palmitic and stearic acids), $> 55\%$ oleic acid, and $< 3\%$ linolenic acid (Lee et al., 2009). These oils would have multiple uses as edible and processed oils (Wilson, 2004). Therefore, it is useful and critical to soybean breeders to have information on the fatty acid composition of new soybean lines.

To date, little has been done on developing high heat tolerant soybean genotypes with high seed quality characteristics under high heat, high humidity, and drought environments such as in the ESPS in the Midsouthern USA. Therefore, the objective of the current research was to evaluate previously developed breeding lines derived from exotic germplasm for yield, germinability, and seed nutritional value under the production environment of the ESPS without relying on irrigation. Further,

we wanted to investigate possible physiological, genotypic, and environmental factors contributing to high germinability under dryland production.

MATERIALS AND METHODS

Description of Experimental Breeding Lines

The seven breeding lines and nine check cultivars evaluated in this study are shown in **Table 1**. Breeding lines 04025-41, 25-1-1-4-1-1, and 34-3-1-2-4-1 were derived from PI 587982A, which was identified by Smith et al. (2008) to have high germinability. The other parent for each of the above three lines was DT98-9102, DT98-9102, and DT97-4290 (Paris et al., 2006), respectively. Breeding line 24-2-1-2-1-2 was derived from DT98-9102 \times PI 603756. The latter PI was also identified by Smith et al. (2008) as having high germinability. Each of the above breeding lines is considered to have 50% exotic parentage. Breeding lines LG03-4561-14 and LG03-4561-19 are sister lines from the same F₂ plant developed by R.L. Nelson and adaptively selected by J.R. Smith for the ESPS. These two lines have 19% exotic parentage derived from PIs 68508 and 445837. Their immediate parents are LG99-5106 and LG97-9226. Breeding line LG04-1459-6 is derived from S32-Z3 \times LG00-3056, but has 25% exotic parentage from PIs 361064, 407710, 189930, and 68600. Two public cultivars from Illinois were included in the study; Dwight (Nickell et al., 1998) and LD00-3309 (Diers et al., 2006). Five commercial cultivars (AG3803, AG3905, AG4403, AG4903, and AG5606) developed by the Monsanto Corporation were included, along with one cultivar developed by Hornbeck Seed Company (C4926) and one cultivar developed by Delta King seed company (DK4866). The genotypes used here were categorized into three groups: (1) breeding lines derived from exotic parental accessions and previously identified to have high germinability under irrigation in the ESPS (04025-41, 25-1-1-4-1-1, 34-3-1-2-4-1, and 24-2-1-2-1-2; all 50% exotic); (2) cultivars (checks); and (3) breeding lines derived from exotic parental accessions and previously identified to have high yield potential under irrigation in the ESPS [(LG03-4561-14 and LG03-4561-19, 19% exotic); (LG04-1459-6, 25% exotic)].

Field Management and Growth Conditions

The experiments were machine planted on 13 April 2012 and 30 April 2013 at the Jamie Whitten Delta States Research Center in Stoneville, MS, USA. The experimental design in each year was a randomized complete block design with three replications. Experimental units were 4-row plots with a 0.91 m row spacing. Plots were 5.79 m long at planting, but end trimmed to 4.88 m long after R1 (beginning bloom; all reproductive stages according to Fehr and Caviness, 1977) and before R6 (full seed-fill). The middle two rows of each plot were harvested with a small plot combine shortly after R8 (full maturity) and weighed as an estimate of seed yield based on 9% moisture. All estimates of seed characteristics (composition, damage, disease, and germinability) were made on this seed for each plot. A field design was implemented so that all plots were accessible for direct combine harvest whenever they were ready for harvest. As such, even

though the maturity groups (MG) ranged from II to V, all plots were timely harvested by the combine as they matured. That is to say, the experiment was not harvested as a group of plots after the last plot matured, but rather over an extended period shortly after each plot arrived at harvest maturity. Timely harvest of each plot was implemented to reduce any potential effect of seed weathering bias. All plots were grown under dry land conditions, with no supplemental irrigation to relieve any drought stress.

Beginning bloom (R1) and full maturity (R8) were recorded for each plot. After R8 and before harvest, plant height and lodging were recorded. Size of harvested seed was estimated as g per 100 seed.

Soil Minerals, N, S, and C Analysis

Nutrients in the soil were analyzed at the University of Georgia's Soil, Plant, and Water Laboratory in Athens, GA. Concentrations of minerals were analyzed on a 5 g soil: 20 ml Mehlich-1 solution and the concentrations were determined using inductively coupled plasma spectrometry (Thermo Jarrell-Ash Model 61E ICP and Thermo Jarrell-Ash Autosampler 300). Soil N, S, and C were determined based on the Pregl-Dumas method (Dumas, 1831; Holmes, 1963; Childs and Henner, 1970) using a C/N/S elemental analyzer having thermal conductivity cells (LECO CNS-2000 elemental analyzer, LECO Corporation, St. Joseph, MI, USA). Briefly, a 0.25 g sample of soil was combusted in an oxygen atmosphere at 1350°C, converting elemental N, S, and C into N₂, SO₂, and CO₂. The gases were then passed through infrared cells and N, S, and C were determined by the elemental analyzer. Average composite random soil samples (four random composite samples across the field), taken at the beginning of the vegetative stage, showed no nutrient deficiency in the soil. The soil texture (clay soil) was in percentage (%) sand = 18, silt = 33.6, and clay = 48.4 with C = 1.4%, N = 0.14%, and organic matter = 1.9%. Nutrient contents were (mg/kg) B = 2.9, Cu = 15.2, Zn = 68.7; and (g/kg) Ca = 5.4, Fe = 21.3, K = 2.7, Mg = 3.5, P = 0.35, and S = 0.21. Leaf analyses did not show any nutrient deficiency (data not shown).

Seed Minerals, N, S, and C Analysis

In all seed analyses, dried seed samples were ground to pass through a 1 mm sieve using a Laboratory Mill 3600 (Perten, Springfield, IL, USA) and ground dried samples were used for analysis. The grinding of all samples in this study was performed under room temperature conditions. Nutrient contents in samples were determined by digesting a 0.6 g ground sample in HNO₃ in a microwave digestion system and nutrients were estimated using inductively coupled plasma spectrometry (Thermo Jarrell-Ash Model 61E ICP and Thermo Jarrell-Ash Autosampler 300; Bellaloui et al., 2011, 2014). Measurements of N, C, and S were conducted on a 0.25 g sample. The samples were combusted, and the percentages of C, N, and S were determined using the C/N/S elemental analyzer (Bellaloui et al., 2011, 2014).

Seed Analysis for Protein, Oil, and Fatty Acids

Dried seed samples were ground to pass through a 1 mm sieve using a Laboratory Mill 3600 (Perten, Springfield, IL, USA) as

TABLE 1 | The 16 soybean genotypes grown under dryland conditions in a 2-year field experiment.

Genotype	Maturity group	Flowering date	R8		Germination		
MS exotic breeding lines ^a		2012	2013	2012	2013	2012	2013
04025-41	IV	Jun 12	Jun 29	Sep 6	Sep 13	57.3	81.7
24-2-1-2-1-2	IV	Jun 6	Jun 22	Sep 1	Sep 6	70.7	91.3
25-1-1-4-1-1	IV	Jun 9	Jun 29	Sep 10	Sep 15	87.3	94.3
34-3-1-2-4-1	IV	May 30	Jun 22	Aug 22	Aug 27	92.0	96.0
CULTIVARS							
AG3803	III	May 20	Jun 5	Aug 9	Aug 24	36.3	50.3
AG3905	III	May 20	Jun 4	Aug 13	Aug 24	55.0	78.7
AG4403	IV	May 20	Jun 9	Aug 15	Aug 29	70.7	78.0
AG4903	IV	May 20	Jun 8	Aug 27	Sep 12	43.0	51.7
AG5606	V	Jun 6	Jun 22	Sep 15	Sep 26	47.3	47.3
C4926	IV	May 31	Jun 22	Sep 5	Sep 13	41.7	54.7
DK4866	IV	May 20	Jun 5	Aug 26	Sep 7	34.0	41.0
Dwight	II	May 18	Jun 3	Aug 9	Aug 20	49.0	58.7
LD00-3309	IV	May 20	Jun 7	Aug 16	Aug 21	51.0	44.3
IL EXOTIC BREEDING LINES^b							
LG03-4561-14	III	May 20	Jun 7	Aug 7	Aug 21	60.3	83.0
LG03-4561-19	III	May 20	Jun 4	Aug 2	Aug 17	65.0	70.0
LG04-1459-6	IV	May 20	Jun 17	Aug 14	Aug 27	80.0	81.7
LSD		1.11	0.57	1.25	1.88	5.1	6.1

^aMS breeding lines, Mississippi breeding lines; ^bIL breeding lines, Illinois breeding lines; R8, full maturity stage. The experiment was conducted in 2012 and 2013 at the Jamie Whitten Delta States Research Center, Stoneville, MS.

described above. Protein, oil, and fatty acids in mature seeds were analyzed according to the detailed methods reported by Bellaloui et al. (2009a, 2010, 2014). Briefly, a 25 g ground sample was analyzed for protein, oil, and fatty acids by near infrared reflectance (Wilcox and Shibles, 2001; Bellaloui et al., 2009a, 2010) using a diode array feed analyzer AD 7200 (Perten, Springfield, IL, USA). The calibration equation was developed by the University of Minnesota using Perten's Thermo Galactic Grams PLS IQ software using conventional chemical protocols with AOAC methods (AOAC, 1990a,b). Protein and oil were expressed on a dry weight basis (Wilcox and Shibles, 2001; Boydak et al., 2002; Bellaloui et al., 2010, 2014). Fatty acid contents (palmitic, stearic, oleic, linoleic, and linolenic acids) were determined on an oil basis (Bellaloui et al., 2009a, 2014).

Boron Determination

The concentration of boron in seeds were measured using the azomethine-H method described by Lohse (1982) and Dordas et al. (2007). Briefly, seed samples were ground to pass through a 1 mm sieve using a Laboratory Mill 3600 (Perten, Springfield, IL, USA) as described above. A ground sample of 1.0 g was ashed at 500°C, extracted with 20 ml of 2 M HCl at 90°C for 10 min, and then a 2 ml sample of the filtered mixture was added to 4 ml of buffer solution (containing 25% ammonium acetate, 1.5% EDTA, and 12.5% acetic acid). A volume of 4 ml of fresh azomethine-H solution (0.45% azomethine-H and 1% of ascorbic acid; John et al., 1975)

was then added. Boron concentration was determined in seeds by a Beckman Coulter DU 800 spectrophotometer (Beckman Coulter, Inc., Brea, CA, USA) at 420 nm (Bellaloui et al., 2014).

Iron Determination

The Fe concentration in seeds was measured according to Bandemer and Schaible (1944) and Loeppert and Inskeep (1996). Seed samples were ground using a Laboratory Mill 3600 (Perten, Springfield, IL, USA) as described above, and samples were then acid digested and extracted, with the resulting reduced ferrous Fe reacting with 1,10-phenanthroline as described by Bellaloui et al. (2011, 2014). Briefly, samples of 2 g of ground sample were acid digested, and the soluble constituents were dissolved in 2 M of HCl. A volume of 4 ml of an aliquot containing 1–20 µg of iron of the sample solution was transferred into a 25 ml volumetric flask and diluted to 5 ml using 0.4 M HCl. A volume of 1 ml of Quinol solution was added to the 5 ml diluted sample solution and mixed. A volume of 3 ml of the phenanthroline solution and 5 ml of the tri-sodium citrate solution (8% w/v) was added. Distilled water was then added to dilute the solution to 25 ml and then incubated at room temperature for 4 h. Phenanthroline reagent solution of 0.25% (w/v) in 25% (v/v) ethanol and quinol solution (1% w/v) was prepared. Concentrations ranging from 0.0 to 4 µg ml⁻¹ of Fe in 0.4 M HCl were prepared for the standard curve. The concentration of Fe was determined by a Beckman Coulter DU 800 spectrophotometer at an absorbance of 510 nm.

Phosphorus Determination

The concentration of P in seed was determined by the yellow phosphor-vanado-molybdate complex method according to Cavell (1955). The detailed description of the method was previously reported by Bellaloui et al. (2009a, 2014). Briefly, seed samples were ground using a Laboratory Mill 3600 (Perten, Springfield, IL, USA) as described above, and a sample of 2 g of ground seed samples was ashed at 500°C, and then 10 ml of 6 M HCl were added. The samples were then placed in a water bath at 100°C until the solution evaporated to dryness. After P was extracted with 2 ml of 36% v/v HCl under heat and filtration, 5 ml of 5M HCl and 5 ml of ammonium molybdate–ammonium metavanadate reagent were added to 5 ml of the filtrate. Ammonium molybdate–ammonium metavanadate was prepared by dissolving 25 g of ammonium molybdate and 1.25 g of ammonium metavanadate in 500 ml of distilled water. A standard curve was produced by preparing a standard solution of phosphorus in a range of concentrations from 0 to 50 µg ml⁻¹ using dihydrogen orthophosphates. The measurement of P concentration was conducted using a Beckman Coulter DU 800 spectrophotometer at an absorbance of 400 nm.

Seed Germination, Seed Vigor, Hard Seed Coat, and Seed Damage Evaluations

Seed assays for standard germination and hard seed were conducted on 200 seeds per plot by the State Seed Testing Laboratory, Mississippi State, MS following the protocol of the Association of Official Seed Analysts (2001). An assay for seed vigor (accelerated-aging germination test) was also conducted by the State Seed Laboratory on a 42 g sample of seed from each plot following standard procedures (Association of Official Seed Analysts, 2002).

Visual ratings of seed coat wrinkling were taken as described by Smith et al. (2008). In short, ratings were taken from seed harvested from each plot as the percentage of visibly wrinkled seed coat surfaces per total visible seed coat surface area. The same visual rating system was used to estimate green seed damage [(Federal Grain Inspection Service (FGIS), 2013)]. Damage ratings (Grain inspection handbook Book II Soybean, 2013) were made for total seed damage for each plot by certified grain inspectors at MidSouth Grain Inspection Service (Stoneville, MS) using a random 125 g sample from each plot.

Fungus Identification and Evaluation

Twenty-five seeds from each plot were disinfected in 0.25% NaOCl for 60 s, blotted dry, plated on acidified potato dextrose agar (APDA; Difco Laboratories, Detroit, MI), and incubated for 7 days at 24°C. *Cercospora kukuchii* was identifiable when cultured on APDA by purple coloration produced on media and the color of the seed coat whereas *P. longicolla* needed further steps to confirm and validate its identity. Identification of *P. longicolla* was based on a single-monoconidial isolate where 10 individual cultures obtained from a monoconidial isolate were evaluated for cultural characteristics (Hobbs et al., 1985; Kulik and Sinclair, 1999; Mengistu et al., 2009, 2007). Each isolate was examined for sporulation, dimensions of conidia, pattern of stroma, and the presence or absence of beta conidia and

perithecia (Hobbs et al., 1985; Barnett and Hunter, 1998; Kulik and Sinclair, 1999).

Experimental Design and Statistical Analysis

The experimental design was a randomized complete block with three replicates (Rep). Analysis of variance was performed using PROC MIXED in SAS [Statistical Analysis System, Copyright 2002–2010, Cary, NC, USA; Windows Version 6.1.7601 (SAS Institute, 2002–2010)]. Year and genotype were modeled as fixed effects. Rep within year was considered as a random effect. Residuals of the random effect factor were shown as covariance parameters in the tables. The residuals refer to Restricted Maximum Residual Likelihood (REML) values, which reflect the total variance of the random parameters in the model. Means were separated by Fisher's protected LSD (0.05). The level of significance of ≤0.05 was used in all measured parameters. Correlation was performed by using Prism (version 6.05) GraphPad Software, La Jolla, CA 92037¹. The correlation was conducted based on the mean values of measured variables.

RESULTS

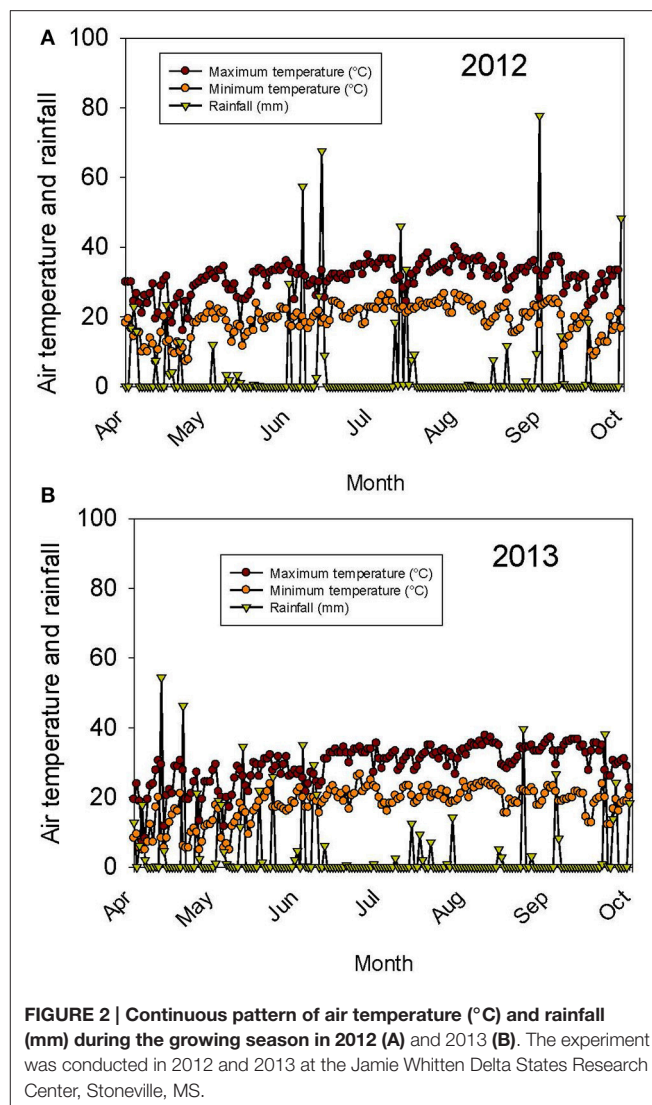
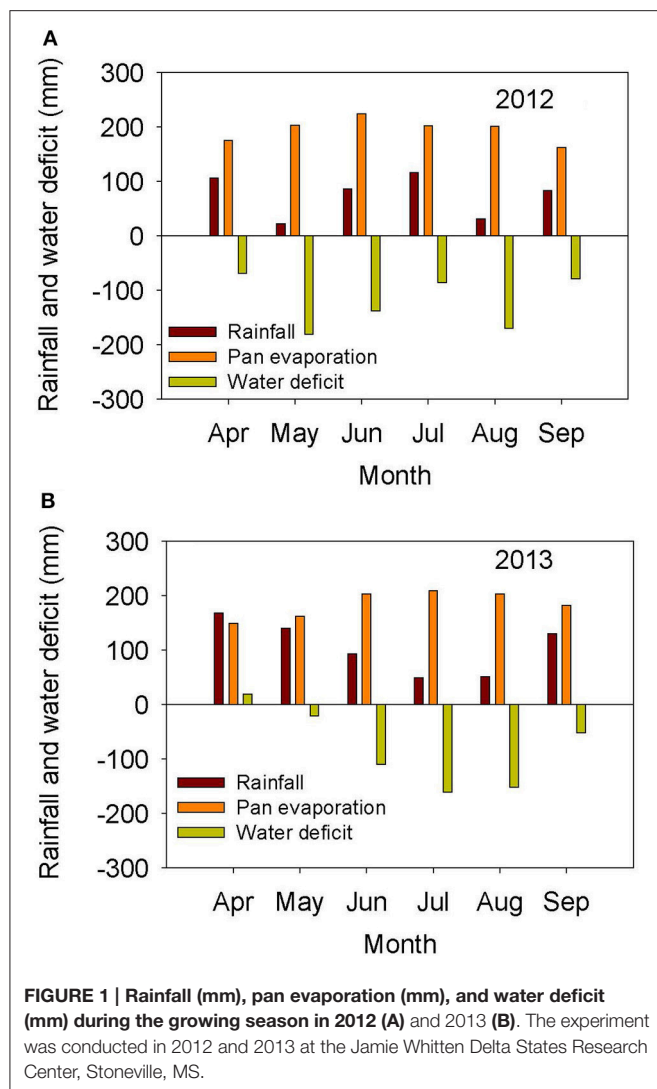
Weather Components

The weather in the Midsouth is characterized by high heat and drought during the growing season in addition to other abiotic and biotic stress factors such as high humidity and diseases such as charcoal rot and phomopsis seed decay. Based on the weather data for 2012 and 2013 (MSUCares, 2016), the month of July in 2013 showed a higher water deficit than in 2012 (−161 mm in 2013 vs. −86 mm in 2012; **Figures 1A,B**). The month of July is an important growing period that coincides generally with the seed-fill stage, especially for early-maturing soybean genotypes. Also, the daily rainfall (**Figures 2A,B**) showed that 2012 received higher rainfall during June and July compared to 2013, but 2013 received more rainfall during August and September than in 2012, indicating differences in rainfall pattern, which may benefit soybean growth and yield for some genotypes as opposed to others due to differences in maturities. The monthly temperature data showed that 2012 was hotter in May through July than in 2013 (**Figures 3A,B**), although the pattern of temperature during the growing season was different (**Figures 2A,B**). In addition, R8 dates among genotypes ranged from 44 days in 2012 to 40 days in 2013, potentially exposing different genotypes to different rainfall and temperature environments in each year. The later planting date in 2013, causing all genotypes to mature later that year, created a potential difference with 2012 for the environment in which each genotype matured.

Analysis of Variance

ANOVA showed that year and genotype were the main source of variability (**Table 2**). Although year × genotype interactions showed significant effects for some seed quality components, the *F*-values were generally low compared with those of main effect factors, except for total seed damage and seed diseases (**Table 2**). The year × genotype interactions are probably due to

¹<http://www.graphpad.com>



yearly changes in rainfall (Figures 1A,B) and heat (Figures 2A,B, 3A,B). No significant effects of year, genotype, and their interaction were observed for *Fusarium*. *Cercospora* was not affected by genotype or year \times genotype interactions. Since year \times genotype was significant for some seed quality components and not for others, we analyzed the data by year to consider the environmental effects (rainfall and temperature).

Seed Germinability, Yield, and Composition (Protein, Oil, and Fatty Acids)

Yield showed significant differences between years. Yield in 2012 was higher than in 2013, except for AG5606 (Table 3), which had an 8.6% increase in 2013. For the other genotypes, the difference in yield between years ranged from a 12.7 to a 58.8% decrease in 2013 over 2012 (Table 3). Two breeding lines (one MS and one IL breeding line) were more stable for yield, with each line having <15% difference in yield between years (Table 3). The MS breeding line, 25-1-1-4-1-1, yielded higher than all other breeding lines and higher than all cultivars except

AG3803, AG3905, and DK4866 in 2013, the more stressful year. Although 2013 was generally cooler than 2012 (Figures 2A,B, 3A,B), rainfall was more uniform during the growing season in 2012 than in 2013 and also more rainfall occurred in July through August (seed-fill stage) in 2012 (Figures 1A,B). It appears that the higher yield in 2012 was likely due to rainfall uniformity and higher rainfall during the seed-fill stage. However, this cannot be generalized, as the higher yield of AG5606 (8.6% higher) in 2013 may be due to its longer maturity (Table 1), which may have better utilized the August and September rains occurring in that year.

Three genotypes (25-1-1-4-1-1, 34-3-1-2-4-1, and LG04-1459-6, had germination rates of $\geq 80\%$ in both years (Table 3). The protein content of the first two genotypes was significantly higher in both years (>40%) than that of all nine cultivars (Table 4). However, LG04-1459-6 had the lowest protein level (36.1% in 2012 and 34.7% in 2013; Table 4) of any genotype tested. This apparent inconsistency between protein and germination is further indicated by the non-significant correlations between

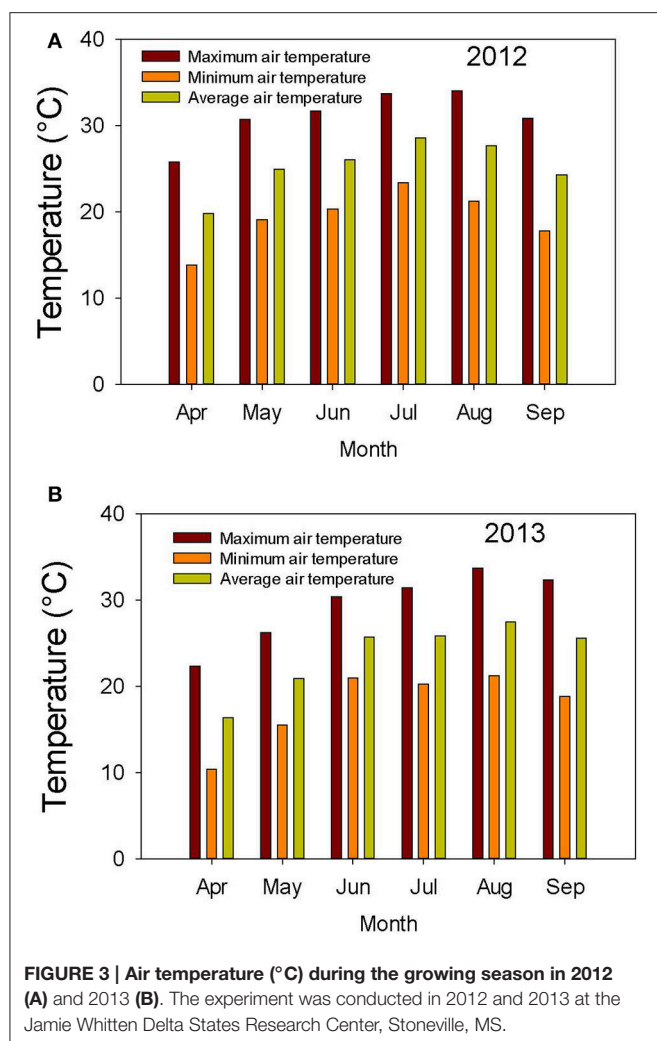


FIGURE 3 | Air temperature (°C) during the growing season in 2012 (A) and 2013 (B). The experiment was conducted in 2012 and 2013 at the Jamie Whitten Delta States Research Center, Stoneville, MS.

germination and protein and oil levels (Table 5). This may indicate that protein level *per se* may not be important for high germination. It may also coincidentally reflect a naturally occurring high protein level in PI 587982A, the parent of both of the above two lines with both high germination and high protein. The association of high germinability and high protein likely needs further study. Among the highest oleic levels in the trial was in AG4403 (32.7% in 2012 and 27.7% in 2013), which had moderate germination percentages of 70.7% in 2012 and 78.0% in 2013. This may indicate that oleic acid level does not have a strong association with germination. Generally, the accumulation of protein and oleic acid was higher in 2012 than in 2013. However, the accumulation of palmitic and linoleic acid was higher only in 2013 (Table 4).

Seed Germinability and Seed Macro- and Micro-Nutrients

Germination ranged from a 13.1% decrease in 2013 over 2012 for LD00-3309 to a 43% increase for AG3905 (Table 3). Except for LD00-3309 and AG5606, germination was higher in 2013 than in 2012. For example, in 2012 only three genotypes (25-1-1-4-1-1,

TABLE 2 | Analysis of variance results for soybean seed yield, seed composition, seed quality, and seed disease measures for 16 soybean genotypes grown under dryland conditions in a 2-year field experiment at Stoneville, MS in 2012 and 2013.

Trait	Year	Genotype	Year x genotype	Residuals
DF	1	15	15	
Seed yield (kg/ha)	11.1*	4.6***	1.84*	2,60,877
100 seed weight (g per100)	19.7*	19.3***	ns	0.57
Protein	63.9***	67.7	2.81***	0.49
Oil	12.4*	39.0***	3.5***	0.21
FATTY ACIDS				
Palmitic	173.8***	10.1***	3.73***	0.07
Stearic	56.2**	6.0***	1.9*	0.01
Oleic	343***	38.1***	5.3***	0.88
Linoleic	251***	30.2***	ns	0.65
Linolenic	31.8**	16.7***	4.7***	0.17
MACRO-NUTRIENTS				
Ca	25.9**	65.3***	5.46***	0.0006
K	13.0*	25.4***	5.9***	0.009
Mg	15.3*	14.2***	3.4***	0.0001
P	105***	24.5***	ns	0.0006
C	995***	13.9***	2.8**	0.11
N	17.5*	31.5***	4.1***	0.025
S	295***	6.7***	ns	0.0002
MICRO-NUTRIENTS				
B	16.1*	70.1***	2.9**	2.62
Cu	71.6***	47.2***	2.7**	0.49
Fe	125***	4.3***	ns	23.46
Mn	21.3**	9.2***	ns	2.53
Mo	10.2*	47.5***	ns	0.52
Zn	127***	10.9***	2.5**	7.52
SEED QUALITY				
Germination	7.4*	21.6***	ns	87.42
Accelerated aging	23.0**	15.5***	3.6***	96.59
Total seed damage	ns	4.5***	6.5***	0.098
Heat damage	ns	ns	ns	0.003
Hard seed	ns	17.2***	2.7**	17.26
Wrinkling	13.5*	14.7***	2.6**	33.19
Green seed	ns	15.0***	4.47***	58.33
SEED DISEASE				
Mold	ns	ns	ns	32.29
Purple seed	4.0*	4.0***	4.0***	1.04
Phomopsis	61.0***	2.5**	1.99*	33.67
Cercospora	10.0**	ns	ns	7.33
Fusarium	ns	ns	ns	61.98
Charcoal rot	ns	1.8*	1.8*	0.49

*Significance at $P \leq 0.05$; **significance at $P \leq 0.01$; ***significance at $P \leq 0.001$.

34-3-1-2-4-1, and LG04-1459-6) had germinations $\geq 80\%$, while in 2013 the number doubled and included 25-1-1-4-1-1, 34-3-1-2-4-1, LG04-1459-6, 04025-41, 24-2-1-2-1-2, and LG03-4561-14. The former three were $\geq 80\%$ in 2012 and in 2013, while the latter three went from germinations of 57, 60, and 71% in 2012 to 82, 83, and 91% in 2013, respectively. The higher germinations for

TABLE 3 | Percentage differences in yield (kg/ha), germination (%), and phomopsis (%) between 2012 and 2013 in each genotype.

Genotype	Yield (kg/ha)			Germination (%)			Phomopsis (%)		
	2012	2013	Diff ^c	2012	2013	Diff	2012	2013	Diff
MS BREEDING LINES^a									
04025-41	2578	1807	771	57.3	81.7	-24.3	20	0	20.0
24-2-1-2-1-2	2679	1585	1094	70.7	91.3	-20.7	28	1.3	26.7
25-1-1-4-1-1	2741	2372	370	87.3	94.3	-7.0	9.3	0	9.3
34-3-1-2-4-1	2602	1991	612	92.0	96.0	-4.0	5.3	0	5.3
CULTIVARS									
AG3803	3687	2865	822	36.3	50.3	-14.0	2.7	0	2.7
AG3905	3797	2670	1128	55.0	78.7	-23.7	12	4	8.0
AG4403	3692	2345	1348	70.7	78.0	-7.3	13.3	1.3	12.0
AG4903	3690	2150	1540	43.0	51.7	-8.7	5.3	0	5.3
AG5606	2114	2295	-182	47.3	47.3	0.0	8	0	8.0
C4926	2509	1722	787	41.7	54.7	-13.0	6.7	0	6.7
DK4866	4122	2428	1695	34.0	41.0	-7.0	14.7	5.3	9.4
Dwight	2374	1766	608	49.0	58.7	-9.7	16	0	16.0
LD00-3309	3405	1403	2002	51.0	44.3	6.7	1.3	0	1.3
IL BREEDING LINES^b									
LG03-4561-14	2571	2246	325	60.3	83.0	-22.7	10.7	2.7	8.0
LG03-4561-19	2990	2096	895	65.0	70.0	-5.0	8	1.3	6.7
LG04-1459-6	3775	2314	1462	80.0	81.7	-1.7	5.3	2.7	2.6
LSD	256	403		5.1	6.1		4.4	1.2	
Min	2114	1403	-182	34.0	41.0	-24.3	1.3	0.0	1.3
Max	4122	2865	2002	92.0	96.0	6.7	28.0	5.3	26.7
Average	3083	2128	955	58.8	68.9	-10.1	10.4	1.2	9.3
Median	2866	2198	858	56.2	74.0	-8.0	8.7	0.0	8.0

^aPercentage exotic of 50%; ^bPercentage exotic ranged from 19 to 25%; ^cDiff, Difference between year 2012 and 2013. The experiment was conducted in 2012 and 2013 at Jamie Whitten Delta States Research Center, Stoneville, MS.

most genotypes in 2013 could be due to the cooler temperature during the seed-fill period, especially between July and early August in 2013 (**Figures 2A,B, 3A,B**). Temperature data showed that the maximum and minimum air temperatures in 2013 were lower than in 2012 during May through July. For example, in June the maximum temperatures were 31.7 vs. 30.4°C, respectively in 2012 and 2013. In July the maximum temperatures were 33.7 vs. 31.4°C, respectively in 2012 and 2013. The same pattern was shown for the minimum temperatures (**Figures 2A,B**). It was interesting that although germination was generally higher in 2013 for most genotypes, yield was generally higher in 2012. For most lines, the environment that produced the highest yield was not the one that produced the highest seed germination.

Seed content of macro-nutrients P and N were significantly higher in two of the genotypes with $\geq 80\%$ germination in both years (25-1-1-4-1-1 and 34-3-1-2-4-1) than in the other genotypes, except for 04025-41 and 24-2-1-2-1-2, which performed similar to the above two high germinability genotypes (**Table 6**). Seed K was significantly higher in 25-1-1-4-1-1 and 34-3-1-2-4-1 than in the rest of the genotypes, except 04025-41, 24-2-1-2-1-2, LG04-1459-6, and AG5606. Calcium content was significantly lower in two of the genotypes with $\geq 80\%$

germination in both years (25-1-1-4-1-1 and 34-3-1-2-4-1). The same was true for 04025-41 and 24-2-1-2-1-2, which performed similar to the above two high germinability genotypes for macro- and micro-nutrient content. There was no consistent trend for Mg, C, and S. Accumulation of Ca, P, C, N, and S in 2012 was almost always higher than in 2013 (**Table 6**). Seed content of micro-nutrients B, Cu, and Mo was significantly higher in two of the genotypes with $\geq 80\%$ germination (25-1-1-4-1-1 and 34-3-1-2-4-1) in both years than in the other genotypes. Again, genotypes 04025-41 (germination rate of 57.3% in 2012 and 81.7% in 2013) and 24-2-1-2-1-2 (germination rate of 70.7% in 2012 and 91.3% in 2013), tended to perform similarly to the above two high germinability genotypes with respect to B, Cu, and Mo. Nutrients Fe, Mn, and Zn had no clear consistent trend between genotypes. The accumulation of B, Cu, Fe, Mo, and Zn was almost always higher in 2013 than in 2012 (**Table 7**). There were significant correlations ($P \leq 0.05$) between germination and Ca, Cu, and K in both years and for P, B, Mo, and N only in 2013 (**Table 5**). The lower nutrients (protein and macro-nutrients) in LG04-1459-6 in spite of its $\geq 80\%$ germination in both years, could be due to the difference of parental lines between LG04-1459-6 in one hand, and 25-1-1-4-1-1 and 34-3-1-2-4-1 on the other hand.

TABLE 4 | Soybean seed composition constituents (protein, oil, and fatty acids, %) of breeding lines and cultivars under dryland conditions.

Genotype	Protein (%)		Oil (%)		Palmitic (%)		Stearic (%)		Oleic (%)		Linoleic (%)		Linolenic (%)	
	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013
MS BREEDING LINES^a														
04025-41	42.5	40.3	21.0	21.6	10.1	11.4	4.2	4.5	33.8	28.1	50.7	53.1	4.3	4.3
24-2-1-2-1-2	42.3	41.2	20.7	21.4	10.6	12.0	4.0	4.4	30.3	23.3	51.4	55.2	4.8	6.6
25-1-1-4-1-1	43.3	40.1	19.3	20.0	10.3	11.6	3.9	4.3	31.8	26.0	51.4	54.4	4.8	4.4
34-3-1-2-4-1	43.6	41.2	20.5	20.1	10.3	11.9	4.3	4.6	32.4	31.4	48.8	51.9	5.6	5.2
CULTIVARS														
AG3803	38.4	38.4	22.6	21.8	10.2	10.8	4.3	4.4	28.2	25.0	53.0	55.9	5.4	6.3
AG3905	38.5	38.1	21.8	20.9	10.3	10.8	4.2	4.4	27.9	26.6	52.5	55.5	5.2	5.4
AG4403	36.9	36.0	23.8	22.8	10.5	11.1	4.4	4.5	32.7	27.7	51.5	53.6	3.6	5.2
AG4903	38.9	37.2	21.3	21.1	9.7	11.0	4.2	4.5	27.8	22.7	54.4	57.4	4.7	6.5
AG5606	39.8	37.7	20.3	20.4	10.1	11.3	3.9	4.5	26.1	21.5	55.1	57.5	6.0	7.3
C4926	37.2	36.5	21.5	21.1	10.3	11.0	4.3	4.6	28.3	23.5	53.9	57.1	4.9	6.3
DK4866	37.3	36.4	21.2	20.9	9.5	10.5	4.1	4.4	27.4	22.7	55.8	57.9	5.4	6.8
Dwight	38.3	37.5	22.3	21.8	10.6	11.1	4.4	4.6	27.0	23.8	53.7	56.2	5.2	6.4
LD00-3309	36.8	37.1	22.6	21.1	10.4	11.2	4.3	4.7	25.8	23.9	54.9	56.5	5.4	5.8
IL BREEDING LINES^b														
LG03-4561-14	35.7	35.4	23.8	22.9	10.1	11.0	4.2	4.5	27.3	23.6	54.2	56.4	5.7	7.1
LG03-4561-19	35.8	35.2	23.8	22.5	10.3	10.6	4.2	4.4	26.7	25.0	53.8	56.0	6.0	6.6
LG04-1459-6	36.1	34.7	23.8	22.8	9.8	10.5	4.1	4.4	27.2	23.8	55.0	57.5	5.6	6.2
LSD	0.33	0.45	0.16	0.33	0.14	0.15	0.06	0.07	0.62	0.42	0.55	0.32	0.28	0.21

^aMS breeding lines, Mississippi breeding lines; ^bIL breeding lines, Illinois breeding lines.

The experiment was conducted in 2012 and 2013 at Jamie Whitten Delta States Research Center, Stoneville, MS.

Seed Germinability, Plant and Seed Physical Characteristics, and Fungal Infection

Beginning bloom (R1), harvest date, and full maturity (R8) were reached later in 2013 than in 2012 (Table 1). This was probably mostly due to the later planting date in 2013 (30 April compared to 13 April in 2012), but may also have been influenced by rainfall and temperature differences between years. For example, rainfall and temperature amounts and patterns were different in each year (Figures 1–3). Even though R8 was reached later in 2013, over all genotypes, the time from R1 to R8 was longer in 2012 than in 2013 (ranging from 1 to 18 days longer with an average of 8 days longer). However, there was no significant ($P > 0.05$) correlation between the time from R1 to R8 and seed yield in either 2012 or 2013. Also, there was no significant ($P > 0.05$) correlation between the time from R1 to R8 and seed germination in 2012 but there was in 2013 ($R = 0.45$; $P = 0.0043$). Cause and effect relationship (regression) between the time from R1 to R8 and seed yield showed no significant effects (Figure 4).

Correlation between Germination and Seed Quality Components

There was a positive correlation between germination and oleic acid and a negative correlation between germination and linoleic acid in both years (Table 5). There was a negative correlation between germination and linolenic acid and a positive correlation

between germination and palmitic acid in 2013 only. A positive correlation between germination and K and Cu was observed in both years, but positive correlations between germination and B and germination and Mo were shown in 2013 only. A negative correlation was shown between germination and Ca and germination and hard seed in both years (Table 5).

DISCUSSION

Seed Germinability, Yield, and Composition

The mechanisms of how seed protein or oleic acid affects germination are still lacking. For example, LeVan et al. (2008) studied the effect of seed composition and seed moisture on germination under a controlled environment and under field conditions and concluded that seed composition may play an important role in imbibitional injury at low seed moisture content. They reported that there was a quadratic relationship between seed protein content and seed germination at different levels of seed moisture, but this relationship was inconclusive for seed oil content and seed germination. They suggested that further research was needed to evaluate the relationship between fatty acids and seed germination. Recently, Chebrolu et al. (2016) studied the effects of three temperature regimes (28/22°C, 36/24°C, and 42/26°C day/night) on heat tolerant line 04025-1-1-4-1-1 (same as 25-1-1-4-1-1 in the current study) and heat-sensitive line DT97-4290 under growth chamber conditions. The germination of seeds of the sensitive genotype had a 50%

TABLE 5 | Correlation between germination and seed quality components in 16 soybean genotypes grown under dryland conditions in a 2-year field experiment.

Trait	2012 R-Value	Significance	2013 R-Value	Significance
Yield	−0.21	ns	−0.01	ns
Protein	0.43	ns	0.42	ns
Oil	−0.07	ns	0.08	ns
Oleic	0.53	*	0.62	*
Palmitic	0.33	ns	0.52	*
Linolenic	−0.07	ns	−0.52	*
Linoleic	−0.63	**	−0.70	**
P	0.49	ns	0.58	*
B	0.37	ns	0.54	*
Cu	0.51	*	0.59	*
Mo	0.46	ns	0.62	**
N	0.42	ns	0.55	*
K	0.53	*	0.54	*
Ca	−0.76	***	−0.61	*
Hard seed	−0.81	***	−0.95	***
Wrinkle	−0.69	**	−0.47	ns
Green seed	−0.34	ns	−0.60	*
Accelerated aging (AA)	0.54	*	0.73	**

*Significance at $P \leq 0.05$; **significance at $P \leq 0.01$; ***significance at $P \leq 0.001$.

The genotypes differed in their germinability. The experiment was conducted in 2012 and 2013 at the Jamie Whitten Delta States Research Center, Stoneville, MS.

reduction at 36°C and was completely inhibited at 42°C. By comparison, the tolerant genotype was unaffected at 36/24°C, but had only 25% germination at 42°C. They also found that the heat sensitive genotype accumulated more seed oil at high temperature (42°C), but it did not differ at 28 or 36°C. Protein content of the heat tolerant genotype, as indicated by nitrogen content, did not differ among temperatures (28, 36, and 42°C). Compared with the heat sensitive genotype, the tolerant genotype accumulated more protein at 36 and 42°C, but no difference was observed between genotypes at 28°C. This may indicate reduction in protein synthesis or protein degradation in the heat sensitive genotype under high heat, compared to the heat tolerant genotype, which maintained similar protein levels at all temperatures. Similarly, 25-1-1-4-1-1 had nearly identical protein levels with AG4903 at four locations in the 2011 Uniform Soybean Test—Southern States, Preliminary Maturity Group IV-S-Late (Gillen and Shelton, 2011), but under the likely more heat stressful conditions of the ESPS, 25-1-1-4-1-1 had significantly higher protein content (43.3 and 40.1% for 2012 and 2013, respectively) than AG4903 (38.9 and 37.2%, respectively) in the current study (Table 4).

In the study of Chebroly et al. (2016), seed oil was higher in the heat sensitive genotype than in the heat tolerant genotype at 42°C, but it was not different between genotypes at 28 or 36°C. They concluded that the higher levels of metabolites such as tocopherols, flavonoids, phenylpropanoids, and ascorbate precursors in the heat tolerant genotype at both temperatures could partially be responsible for the observed heat tolerance

in 25-1-1-4-1-1. Our results are partially in agreement, as oil was not correlated with germination, and the daily high temperature in the field was about 36°C, but did not reach 42°C as in the study of Chebroly et al. (2016). Further research is needed to evaluate the relationship between fatty acids and seed germination, as the relation between oil and germination is still inconsistent (LeVan et al., 2008). The change in seed composition constituents in 2012 and 2013 is in agreement with previous research indicating that temperature was considered a contributing factor to the variability of seed composition, and the increase or decrease of seed oil or protein concentration was associated with the range of temperatures under which soybean seeds mature (Bellaloui et al., 2009a). In our study, 2013 was drier than 2012, especially in July (coinciding with seed-fill for the early maturing genotypes (Figures 1A,B). Furthermore, 2012 was warmer than 2013 (Figures 2A,B). Therefore, the patterns of both rainfall and temperature were different in each year (Figures 2A,B, 3A,B). Hence, it is likely that most of the changes in seed constituents between years were due to drought (Lamoine et al., 2013; Andrade et al., 2016) and heat (LeVan et al., 2008; Chebroly et al., 2016).

Seed Germinability and Seed Macro- and Micro-Nutrients

Previous research showed that PI 587982A and PI 603723 (genotypes with high seed germination under high temperatures) had higher germination rates and lower hard seed compared to the poor seed germination genotypes PI 80480, PI 84976-1, DSR-3100 RR STS, and Pella 86 (Bellaloui et al., 2012b). Orazaly et al. (2014) observed that Ca content in the seed coat had an effect on water absorption. Other studies found Ca to be negatively correlated with water absorption (Saio et al., 1973; Saio, 1976) and positively correlated with hard seed (Zhang et al., 2009), and that low water absorbance resulted in hard seed and low germination rate. Low Ca and hard seed are important because they affect the texture of soybean natto (Mullin and Xu, 2001), are positively correlated with cooked seed hardness, and are influenced by environmental factors such as temperature and soil type (Chen et al., 2001). The role of Ca in hard seed resides in the important role of Ca in the cell membrane and its contribution to cell wall thickness due to a calcium oxalate cell layer in the cell wall (Webb, 1999). In the current study, Ca was found to be significantly ($P < 0.05$) negatively correlated with germination in both years ($R = -0.76$ and -0.61 in 2012 and 2013, respectively; Table 5). However, Ca was significantly ($P < 0.01$) positively correlated with hardseededness ($R = 0.81$ and 0.68 in 2012 and 2013, respectively).

The significantly higher K, P, N, B, Cu, and Mo content in seeds of 25-1-1-4-1-1 and 34-3-1-2-4-1, compared to all other genotypes with $\geq 80\%$ germination in both years, indicates that these nutrients might be important for germination. If this is true, maintaining higher levels of these nutrients in soil and seed could contribute to germination and overall seed health. Bishnoi et al. (2007) studied the effects of Ca and P on soybean seed production and quality and found that the application of Ca at 100 and P at 90 kg/ha significantly improved soybean seed yield

TABLE 6 | Soybean seed macro-nutrients (%) in breeding lines and cultivars under dryland conditions.

Genotype	Ca (%)		K (%)		Mg (%)		P (%)		C (%)		N (%)		S (%)	
	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013
MS BREEDING LINES														
04025-41	0.26	0.25	2.32	2.27	0.20	0.24	0.60	0.65	49.0	53.1	6.35	6.52	0.26	0.33
24-2-1-2-1-2	0.21	0.36	2.24	2.07	0.23	0.26	0.57	0.64	49.0	52.9	6.39	6.40	0.28	0.32
25-1-1-4-1-1	0.23	0.27	2.25	2.08	0.21	0.24	0.63	0.63	48.4	52.5	6.21	6.56	0.26	0.33
34-3-1-2-4-1	0.28	0.36	2.37	2.15	0.23	0.25	0.59	0.64	49.1	52.4	6.37	6.57	0.29	0.33
CULTIVARS														
AG3803	0.42	0.40	1.70	1.95	0.24	0.24	0.48	0.52	50.0	53.6	5.64	5.27	0.27	0.33
AG3905	0.40	0.47	1.57	1.86	0.25	0.27	0.50	0.56	49.6	52.9	5.70	5.50	0.26	0.32
AG4403	0.38	0.44	1.47	1.85	0.23	0.26	0.46	0.55	50.3	53.3	5.63	5.71	0.24	0.31
AG4903	0.43	0.43	1.88	1.91	0.25	0.25	0.49	0.56	49.2	52.8	5.75	6.00	0.26	0.33
AG5606	0.39	0.36	1.85	2.05	0.23	0.25	0.51	0.57	49.0	52.8	5.87	5.65	0.26	0.31
C4926	0.41	0.42	1.73	1.81	0.21	0.23	0.50	0.56	49.1	53.1	5.51	5.76	0.25	0.31
DK4866	0.49	0.50	1.70	1.92	0.28	0.27	0.47	0.53	49.2	52.7	5.52	5.75	0.23	0.29
Dwight	0.42	0.46	1.71	1.88	0.28	0.27	0.51	0.53	49.9	53.1	5.65	5.98	0.25	0.30
LD00-3309	0.50	0.56	1.66	1.93	0.29	0.27	0.49	0.55	49.5	52.3	5.49	6.06	0.24	0.30
IL BREEDING LINES														
LG03-4561-14	0.38	0.42	1.93	2.05	0.26	0.26	0.52	0.53	50.4	53.5	5.35	5.76	0.26	0.31
LG03-4561-19	0.34	0.40	1.83	2.04	0.26	0.25	0.51	0.55	50.6	53.4	5.34	5.68	0.25	0.30
LG04-1459-6	0.32	0.37	1.75	2.05	0.25	0.25	0.44	0.51	50.6	53.6	5.33	5.80	0.23	0.29
LSD	0.02	0.01	0.06	0.05	0.01	0.01	0.01	0.01	0.20	0.19	0.08	0.10	0.01	0.01

MS breeding lines, Mississippi breeding lines; IL breeding lines, Illinois breeding lines.

The experiment was conducted in 2012 and 2013 at the Jamie Whitten Delta States Research Center, Stoneville, MS.

and quality (germination and vigor), but seed quality decreased due to weathering and diseases if harvest was delayed (Bellaloui et al., 2012a). Bishnoi et al. (2007) also found that seed quality (viability and germination %) was enhanced with application of Ca at 100 and P at 90 kg/ha in comparison to plots without application. The positive response of soybean to Ca application in the study of Bishnoi et al. (2007) could be due to Ca deficiency in the soil or reduced Ca supply, as reduced Ca supply to the plant may reduce seed Ca concentration, resulting in poorer seed germination (Keiser and Mullen, 1993). The genotypes used in the current study were different from those used by Bishnoi et al. (2007), which may partially explain the apparent differences in the effect of Ca between the current study and that of Bishnoi et al. (2007).

The significant role of macro-nutrients such as N, P, K, S, and Ca and micro-nutrients such as Fe, Zn, and B to plant growth, development, yield, and quality has been well-documented (Mengel and Kirkby, 1982; Marschner, 1995; Samarah and Mullen, 2004). For example, Haq and Mallarino (2005) reported that N, P, K, and other nutrients can affect several physiological processes that, in turn, could affect grain yield and protein or oil content. Different levels of seed nutrients among genotypes could be due to genotypic background differences and to their different responses to environment, especially drought and heat. It was previously reported that the content of seed micro- and macro-nutrients was found to be influenced by environment and genotype (Zhang et al., 1996; Haq and Mallarino, 2005; Bellaloui et al., 2011, 2015).

Seed Germinability, Plant and Seed Physical Characteristics, and Fungal Infection

Seed physical characteristics (quality) and seed diseases are shown in **Tables 8, 9**, respectively. The response of genotypes for hard seed differed in each year, indicating the contribution of both genotypic and environmental factors to this trait. In terms of environment, 2013 produced a higher level of hardseededness than 2012 for most genotypes. In terms of genotypic effect, the four breeding lines derived from PIs 587982A and 603756 (the MS breeding lines) averaged <1% hard seed in both 2012 and 2013, whereas the level of hard seed of the nine cultivars ranged from 1.3 to 17.0% in 2012 and from 4.0 to 27.7% in 2013 (**Table 8**). Clearly, there was a major genotypic effect. Recently, a major single recessive gene (*isc*) for permeable seed coat was identified in PI 587982A (Kebede et al., 2014), whose permeable seed coat effect can be observed in the PI 587982A-derived lines 04025-41, 25-1-1-4-1-1, and 34-3-1-2-4-1 (**Table 8**).

Total seed damage is the official total measure of grain damage as prescribed by the United States Federal Grain Inspection Service (FGIS). It includes grain damage due to multiple factors, including mold, heat, green seed, stink bug, etc. Grain elevators assess discounts on the value of grain produced by soybean producers based on FGIS standards. This can result in a loss of revenue to producers when they sell their grain. A common level of grain damage that could result in discounting at grain elevators is the 2% level, meaning that damage >2% would result

TABLE 7 | Soybean seed micro-nutrient concentrations (mg/kg) in breeding lines and cultivars under two dryland environments.

Genotype	B (mg/kg)		Cu (mg/kg)		Fe (mg/kg)		Mn (mg/kg)		Mo (mg/kg)		Zn (mg/kg)	
	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013
MS BREEDING LINES^a												
04025-41	41.2	44.0	19.5	19.9	65.5	87.3	21.2	21.9	7.4	9.9	52.0	60.2
24-2-1-2-1-2	42.5	40.9	19.5	20.5	67.3	83.6	19.6	21.9	6.5	9.1	56.6	63.3
25-1-1-4-1-1	41.9	41.3	20.3	20.1	67.4	81.7	16.6	18.5	9.4	10.6	57.6	61.8
34-3-1-2-4-1	42.8	42.1	18.5	20.1	69.1	88.3	21.9	24.6	7.5	9.5	56.1	61.1
CULTIVARS												
AG3803	32.7	37.2	13.4	15.7	60.9	75.1	21.1	25.2	3.7	4.5	42.9	53.1
AG3905	28.3	33.2	14.9	17.9	69.2	85.5	21.5	26.9	3.6	5.0	48.7	61.0
AG4403	27.1	31.9	13.2	16.2	59.1	81.2	19.3	22.5	3.5	4.7	45.3	60.2
AG4903	30.3	31.7	14.7	17.3	67.5	81.7	21.0	23.3	4.5	5.2	50.3	62.9
AG5606	33.0	33.7	17.1	18.3	70.1	88.8	18.5	20.9	6.1	6.1	48.5	56.3
C4926	38.5	39.8	14.5	16.5	65.3	77.1	21.3	22.7	4.7	5.9	51.2	60.9
DK4866	30.0	31.5	14.8	17.2	66.1	82.0	20.7	24.3	4.4	5.1	50.1	61.1
Dwight	26.9	28.1	14.9	16.0	76.1	88.8	24.0	23.9	3.0	4.4	50.2	63.9
LD00-3309	25.9	30.4	15.3	16.9	70.1	82.5	23.4	25.9	3.1	4.7	48.3	65.3
IL BREEDING LINES^b												
LG03-4561-14	33.2	36.6	15.1	16.3	67.6	81.4	20.6	21.1	4.1	5.0	47.5	57.0
LG03-4561-19	30.4	36.1	14.5	15.6	67.3	89.0	19.9	20.6	3.5	4.8	45.9	59.1
LG04-1459-6	25.9	29.3	14.6	17.4	58.9	75.9	17.4	20.7	3.1	4.5	41.3	53.4
LSD	1.04	0.86	0.35	0.44	2.80	2.79	1.03	0.80	0.44	0.52	1.47	1.74

^aMS breeding lines, Mississippi breeding lines; ^bIL breeding lines, Illinois breeding lines.

The experiment was conducted in 2012 and 2013 at Jamie Whitten Delta States Research Center, Stoneville, MS.

in discounting of payments to producers. In the current study, one Illinois-derived breeding line (LG04-1459-6) had damage >2% (2.1%) in 1 year (2013; **Table 8**). All other lines had total grain damage of <2% in 2012 and 2013 and so would not have been assessed damage charges under the conditions of this study.

Seed coat wrinkling is a type of seed damage that is measured by FGIS standards, but is not discounted by elevators. Even so, it has been shown to be negatively correlated with seed germination and seed vigor (Smith et al., 2008). As with many physical seed characteristics, it is influenced by both environment and genotype. In the current study, most genotypes showed higher wrinkling in 2012 than in 2013 (**Table 8**), indicating an effect due to environment. Recent studies have also shown a genetic effect involved in the level of seed coat wrinkling. Kebede et al. (2013) identified a major single dominant gene (*Wri*) in PI 567743 that controls the level of seed coat wrinkling observed in high heat environments, such as the ESPS. The two MS breeding lines (25-1-1-4-1-1 and 34-3-1-2-4-1) derived from PI 587982A had significantly lower levels of seed coat wrinkling than all other lines tested in both years. For example, cultivar C4926 had wrinkling scores of 50 and 40% in 2012 and 2013, respectively, whereas 25-1-1-4-1-1 had wrinkling scores of 13.3 and 3.3% in those years, respectively (**Table 8**). Over all genotypes there was a significant negative correlation between germination and wrinkling in 2012 ($R = -0.69$; $P = 0.01$)

but not in 2013 ($P > 0.05$) the year with less wrinkling (**Table 5**).

Green seed damage is assessed by FGIS standards, but in that system there must be a minimum intensity of green before it is reported as damage. Generally, an intensity of light green is not reported as damaged by FGIS standards. The green seed damage estimates for the current study reported any level of green observed. Light green and dark green shades were recorded equally as green seed damage. Green seed damage in soybean is known to be caused by rapid dry down of maturing seed, where the normal degradation of chlorophyll is inhibited (Adams et al., 1983). Hence, any stress (drought, hard freeze, and high heat) that does not allow for the normal slow dehydration and chlorophyll degradation of the seed will promote green seed damage. Rapid dry down of maturing seed, and its resulting green seed damage, is harmful to soybean germination (Green et al., 1965) because it does not allow for the production of the germination-specific enzymes malate synthase and isocitrate lyase (Adams et al., 1983). Hence, environment has a large effect on the level of green seed damage (Green et al., 1965; Adams et al., 1983). In spite of the large potential effect of environment, there was no year effect in the current study (**Table 2**). Yet, there was a significant year \times genotype interaction (**Table 2**), indicating that some genotypes responded differently to specific within-year environments. This is understandable, given that the 16 genotypes ranged in maturity from MG II to V and experienced maturation under different environmental conditions. For example, MG III AG3905 had

50% green seed damage in 2012, but 26.7% green seed damage in 2013 (Table 8). The earlier planting date (13 April) affected an earlier harvest in 2012 and may have promoted higher green seed damage, as suggested in the study of Green et al. (1965). However, MG IV C4926 matured during different environments in both years and had only 3.3% green seed damage in 2012, but 26.7% in 2013 (Table 8). Green seed damage had a highly significant genotype component in this study (Table 2), indicating that genotypes do not respond the same to the same stress. This can be observed in comparisons between AG3803 and AG3905. Both had similar maturation (Table 1) environments in each year, but AG3803 had no green seed, while AG3905 had high levels, as noted above (Table 8). Likewise, high-germination breeding line 34-3-1-2-4-1 had no green seed damage in either year (Table 8). High-germination breeding line 25-1-1-4-1-1 had 0 and 10% green seed damage in 2012 and 2013, respectively, compared to 3.3 and 26.7% green seed damage for C4926 in 2012 and 2013, respectively (Table 8). Over all genotypes there was not a significant ($P > 0.05$) negative correlation between germination and green seed damage in 2012 but there was in 2013 ($R = -0.60$; $P < 0.05$), the more stressful year (Table 5). More research is needed on the inheritance of tolerance to green seed damage in soybean.

The infection for *C. kukuchii* and *Fusarium* spp. was generally low (Table 9). An important exception was the significantly higher level of *Fusarium* spp. observed on 25-1-1-4-1-1 (25.3%) in 2012 (Table 9). This moderately high level of infection, together with the 9% *P. longicolla* infection (Tables 3, 9) may have negatively affected the percent germination of 25-1-1-4-1-1 (87.3%) in 2012 (Table 3), as the germination of 25-1-1-4-1-1 was 94.3 in 2013, when its level of *Fusarium* was only 2.7% and its level of *P. longicolla* was zero (Table 9). It may seem surprising that these higher levels of fungal infection appeared to have had an insignificant effect on the FGIS total seed damage of 25-1-1-4-1-1, which was 0.77 in 2012 and 0.23 in 2013 (Table 8). But it is frequently observed that seed which appears to have no visible sign of disease will be found to be infested with *P. longicolla* or *Fusarium* spp when plated onto media. FGIS damage ratings are totally visual.

Phomopsis longicolla infection (Tables 3, 9) was higher in 2012 than in 2013 and significant differences were found among the lines for *P. longicolla* infestation in 2012. The higher infection in 2012 is most likely due to higher rainfall in 2012 during seed-fill (from August through mid-September), as well as a more positive rain distribution pattern across the growing season (Figures 1, 2). Though germination and *P. longicolla* levels were not significantly correlated (data not shown) it was observed that the 2012 germination of breeding lines 04025-412 and 24-2-1-2-1-2 (57.3 and 70.7%, respectively) were probably negatively impacted by their high *P. longicolla* levels (20 and 28%, respectively) in 2012 as their germination levels in 2013 were much higher (81.7 and 91.3%, respectively).

In some US states, such as Mississippi, the minimum germination for certification is 80% (Keith and Delouche, 1999). Therefore we further evaluated traits based on germination levels of at least 80% within a given environment (year). Comparison of quality components between genotypes with $\geq 80\%$ and

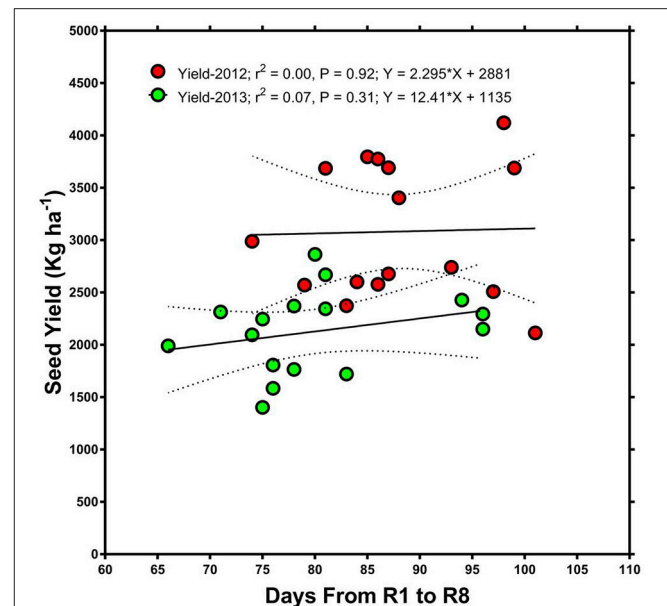


FIGURE 4 | Relationship between days after planting (time from beginning of flowering, R1, to full maturity stage, R8) and seed yield. The graph shows that there was no significant relationship between the time from R1 to R8 and seed yield in either 2012 or 2013.

those with $<80\%$ germination identified differences between groups for germination, AA, and Ca in both 2012 and 2013, whereas flowering date, hardseededness, seed coat wrinkling, palmitic acid, N, P, K, B, Cu, and Mo were only significant in 1 year (all in 2013; Table 10). Those traits significantly different only 2013 were probably a result of weather differences between years. The lower Ca was associated with the high-germination group in both years is likely very meaningful (Tables 5, 10) and should be investigated in greater detail. Also of interest is what variables were not different between germination groups in this comparison. Differences in seed yield, seed size, maturity, and total seed damage were not significant in either year between high and low germination groups. This may indicate that genotypes with high germinability could be either low yielding or high yielding, and could be of multiple maturities. Hence, selection for high yield with high germination, and of multiple maturities, should be possible.

Previous research on seed germination, accelerated aging, hard seed, wrinkling, and diseases (phomopsis and charcoal rot) showed that the high germinating genotypes had the lowest hard seed and seed wrinkling percentages (Smith et al., 2008; Mengistu et al., 2009, 2010). For example, Smith et al. (2008) evaluated seed quality characteristics for 513 soybean lines (486 accessions, 24 ancestors, and cultivars Stalwart, Croton 3.9, and Stressland) with maturity groups ranging from 000 to MG V under field conditions in the early soybean production system at Stoneville, MS, in 2002 and 2003. They found significant ($P = 0.01$) negative correlations between standard germination and hard seed, wrinkled seed, phomopsis, and seed weight. Similar results

TABLE 8 | Soybean seed quality components (%) in breeding lines and cultivars under two dryland environments.

Genotype	Wrinkling		Accelerated aging		Total seed damage		Hard seed		Green seed	
	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013
MS BREEDING LINES^a										
04025-41	30.0	13.3	26.3	54.3	0.53	0.40	0.0	0.7	0.0	0.0
24-2-1-2-1-2	20.0	10.0	71.7	69.3	0.33	0.27	0.3	0.0	0.0	10.0
25-1-1-4-1-1	13.3	3.3	82.7	90.3	0.77	0.23	0.7	0.3	0.0	0.0
34-3-1-2-4-1	10.0	0.0	87.3	94.7	0.30	0.23	0.0	0.0	0.0	0.0
CULTIVARS										
AG3803	30.0	26.7	19.3	55.7	0.27	0.40	12.0	19.7	50.0	26.7
AG3905	23.3	26.7	38.3	66.3	0.47	0.20	2.7	4.7	20.0	6.7
AG4403	20.0	23.3	39.0	69.7	0.13	0.27	1.3	4.0	23.3	10.0
AG4903	30.0	23.3	38.3	54.0	0.50	0.33	13.0	16.3	10.0	16.7
AG5606	30.0	20.0	30.3	32.3	0.43	0.23	10.3	12.3	3.3	26.7
C4926	50.0	40.0	51.0	68.3	0.70	0.27	14.0	7.3	16.7	33.3
DK4866	30.0	20.0	64.3	55.3	0.60	1.37	5.3	22.0	0.0	20.0
Dwight	26.7	20.0	49.3	69.7	1.30	0.10	15.3	10.7	20.0	6.7
LD00-3309	16.7	10.0	59.3	62.0	0.47	0.13	17.0	27.7	26.7	23.3
IL BREEDING LINES^b										
LG03-4561-14	20.0	26.7	37.0	79.0	0.30	0.13	0.3	2.3	40.0	30.0
LG03-4561-19	16.7	26.7	58.7	70.0	0.53	0.20	6.0	8.3	26.7	33.3
LG04-1459-6	26.7	16.7	39.3	75.7	0.23	2.10	0.7	5.0	10.0	13.3
LSD	3.1	3.4	5.8	5.9	0.18	0.20	1.7	3.0	3.6	4.8

^aMS breeding lines, Mississippi breeding lines; ^bIL breeding lines, Illinois breeding lines.

The experiment was conducted in 2012 and 2013 at Jamie Whitten Delta States Research Center, Stoneville, MS.

were found for seed germination and hard seed by Bellaloui et al. (2012b). Also, it was found that *Phomopsis longicolla* Hobbs caused substandard germination (Mengistu and Heatherly, 2006; Smith et al., 2008; TeKrony et al., 1984), and high temperature with wet and dry conditions increased seed coat wrinkling, reducing seed germination (Franca-Neto et al., 1993).

Mengistu et al. (2009) evaluated soybean genotypes of different maturities for seed quality characteristics (seed germination, seed phomopsis infection, hard seed) under different irrigation regimes (non-irrigated, irrigated pre-flowering, and irrigated after flowering). They found that soybean genotypes with higher germination rates had lower phomopsis seed infection, lower hard seed rate, and the germination rate and seed phomopsis infection depended on irrigation type. Therefore, as already noted above, the effect of seed diseases on germination appears to be dependent on genotypic response to the pathogen, severity and threshold of infection, and environmental factors (drought and temperature) and their interactions. Therefore, further research including a larger number of genotypes with higher levels of infestation may show different results.

The causes of poor seed quality in the ESPS were suggested to be related to temperature, soil moisture, and disease infection during the periods from the beginning of seed-fill to full maturity and pre-harvesting, leading to hard seed and low seed viability and vigor (TeKrony et al., 1980; Roy et al., 1994). The differences in hard seed of soybean genotypes have been attributed to genetic

variation (Kilen and Hartwig, 1978; Kebede et al., 2014) and environmental conditions (Smith et al., 2008; Mengistu et al., 2009), and may play an important role in preventing phomopsis infection. Hard seed can be considered an undesirable trait, as it lowers seed germination and negatively impacts processing soybean to soy food, which leads to poor quality and adverse cost factors (Mullin and Xu, 2001). The hard seed trait is related to moisture impermeability and seed coat character. The hard seed trait has also been considered a positive trait. Roy et al. (1994) evaluated seed infection in soybean cultivar Forrest (susceptible to infection and with permeable seed coat) and D67-5677-1 (a breeding line with impermeable seed coat). They found that after injecting *Phomopsis longicolla* conidia into the seed cavities of pods that high levels of seed infection occurred in Forrest, but not in the hard seed genotype D67-5677-1. They reported that genotypes D67-5677-1, D86-4629, D86-4565, and D86-4669 (hard seed lines) all expressed resistance to naturally occurring infection by *Phomopsis* during 8 years of evaluations (Roy et al., 1994). They also indicated that impermeable seeds and level of seed infection with *Phomopsis* were negatively correlated. They concluded that although seed coat impermeability *per se* conferred resistance to phomopsis, other research suggested that impermeability alone did not account for the resistance, as much lower correlations were obtained between seed coat impermeability and phomopsis seed infection when a larger number of hard seed genotypes were used (Roy et al., 1994). Clearly, further research is needed to determine the relative

TABLE 9 | Soybean seed diseases (infection %) in breeding lines and cultivars under two dryland environments.

Genotype	Mold (%)		Purple seed (%)		Phomopsis (%)		Cercospora (%)		Fusarium (%)		Charcoal rot (%)	
	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013	2012	2013
MS BREEDING LINE^a												
04025-41	0.0	3.3	0.0	0.0	20.0	0.0	4.0	0.0	4.0	0.0	0.0	0.0
24-2-1-2-1-2	0.0	0.0	0.0	0.0	28.0	1.3	1.3	0.0	5.3	0.0	0.0	0.0
25-1-1-4-1-1	0.0	0.0	0.0	0.0	9.3	0.0	4.0	1.3	25.3	2.7	0.0	0.0
34-3-1-2-4-1	0.0	0.0	0.0	0.0	5.3	0.0	0.0	0.0	1.3	0.0	0.0	0.0
CULTIVARS												
AG3803	0.0	0.0	0.0	0.0	2.7	0.0	0.0	1.3	1.3	0.0	0.0	0.0
AG3905	0.0	0.0	0.0	0.0	12.0	4.0	0.0	0.0	0.0	0.0	0.0	0.0
AG4403	10.0	0.0	0.0	0.0	13.3	1.3	5.3	2.7	0.0	0.0	0.0	0.0
AG4903	0.0	3.3	0.0	0.0	5.3	0.0	4.0	1.3	1.3	0.0	0.0	0.0
AG5606	0.0	3.3	0.0	0.0	8.0	0.0	4.0	0.0	9.3	1.3	0.0	0.0
C4926	0.0	3.3	0.0	0.0	6.7	0.0	6.7	0.0	0.0	2.7	0.0	0.0
DK4866	0.0	3.3	6.7	0.0	14.7	5.3	8.0	0.0	2.7	1.3	0.0	0.0
Dwight	0.0	13.3	0.0	0.0	16.0	0.0	2.7	1.3	1.3	0.0	1.3	0.0
LD00-3309	0.0	0.0	0.0	0.0	1.3	0.0	0.0	0.0	0.0	1.3	0.0	0.0
IL BREEDING LINES^b												
LG03-4561-14	0.0	0.0	0.0	0.0	10.7	2.7	0.0	1.3	0.0	0.0	1.3	0.0
LG03-4561-19	0.0	0.0	0.0	0.0	8.0	1.3	0.0	0.0	0.0	0.0	2.7	0.0
LG04-1459-6	0.0	3.3	0.0	0.0	5.3	2.7	0.0	2.7	0.0	1.3	0.0	0.0
LSD	2.4	3.8	0.8	0.0	4.4	1.2	2.0	0.9	6.1	1.1	0.6	0.0

^aMS breeding lines, Mississippi breeding lines; ^bIL breeding lines, Illinois breeding lines.

The experiment was conducted in 2012 and 2013 at Jamie Whitten Delta States Research Center, Stoneville, MS.

contribution of the hard seed trait to disease resistance and seed quality using a larger number of genotypes with more severe disease infection under the stress environmental factors of high heat and water deficit conditions.

Correlation between Germination and Seed Quality Components

Previous research reported the possible involvement of seed composition constituent levels in germination. For example, LeVan et al. (2008) suggested a quadratic relationship between seed protein content and standard seed germination, and found that the variability in seed protein content did not change the quadratic relationship between seed protein content and seed germination. However, they also found that the relationship between seed oil content and seed germination was not conclusive, which was partially supported by our results in that there was no correlation found between germination and protein or oil. Previous research on minerals showed that high germinability breeding lines correlated with seed soluble and structural B (Bellaloui et al., 2008). In this work, B was positively associated with germination in 2013 ($r = 0.54$), but not in 2012 (Table 5). Calcium content in the seed coat was found to be positively correlated with hard seed (Zhang et al., 2009), and influenced by environmental factors such as temperature and soil type (Chen et al., 2001). Smith et al. (2008) evaluated 513 soybean lines and found a significant ($P = 0.01$)

negative correlation between standard germination and hard seed ($R = -0.40$), wrinkled seed ($R = -0.53$), phomopsis seed infection ($R = -0.56$), and seed weight ($R = -0.21$), in agreement with our finding for hard seed and wrinkled seed, but not for phomopsis seed infection and seed weight. The lack of correlation of disease infection and seed weight with germination in our experiment could be due to genotype differences as well as environmental factors, especially heat and drought, which may have reduced the incidence of disease. Soybean lines in the current experiment were grown under dryland conditions, differing from those of Smith et al. (2008), where irrigation was applied to alleviate water stress. The severity of diseases such as *Phomopsis* infection to plants was reported to be influenced by environment, involved altered seed composition constituents (Bradley et al., 2002; Bellaloui et al., 2012b), and was affected by temperature, rain/irrigation, genotype, and crop management (Mengistu et al., 2010; Bellaloui et al., 2012a). Research available on the relationship between germination and seed composition constituents is still limited and further research is needed to evaluate the relationship between fatty acids and seed germination (LeVan et al., 2008).

Overall Important Discussion Points

This study involved soybean genotypes from three distinct germplasm pools; the nine cultivars typify the current pool available to producers, the three breeding lines from the

TABLE 10 | Difference between lines with <80% germination and those with < 80% germination for plant and seed characteristics.

Trait	2012		2013	
	Difference*	P-Value	Difference	P-Value
AGRONOMY/PHYSIOLOGY MEASURES				
Seed yield (kg/ha)	53.33	0.9012	121.57	0.5700
100 Seed weight (g per100 seed)	1.15	0.2457	1.01	0.1339
Flowering date	-5.03	0.3929	-10.48	0.0288
R8	-4.47	0.6243	-2.06	0.7476
Harvest date	-4.99	0.6162	-1.16	0.8580
SEED QUALITY (%)				
Germination	-34.03	0.0004	-30.53	0.0002
Accelerated aging (AA)	-24.93	0.0443	-16.89	0.0233
HS (Hard seed)	13.37	0.0571	22.68	0.0004
Total seed damage (DKT)	0.07	0.6984	-0.21	0.4623
Green seed	14.87	0.1358	11.44	0.0588
Wrinkle	9.74	0.1080	12.00	0.0149
Stinkbug damage	-0.02	0.8037	0.00	0.9362
Mold	0.77	0.6472	1.56	0.4287
PROTEIN AND OIL (%)				
Protein	-2.64	0.1294	-1.81	0.0913
Oil	0.83	0.3813	-0.03	0.9535
FATTY ACIDS (% OF OIL)				
Palmitic	0.06	0.7647	-0.49	0.0321
Stearic	0.10	0.3161	0.05	0.2964
Oleic	-2.05	0.2206	-1.77	0.1801
Linoleic	1.69	0.1759	1.59	0.0719
Linolenic	-0.23	0.5940	0.62	0.1870
MACRO-NUTRIENTS (%)				
C	0.21	0.6469	-0.02	0.9452
Ca	0.11	0.0381	0.11	0.0042
K	-0.31	0.0826	-0.19	0.0004
Mg	0.02	0.2379	0.01	0.2772
N	-0.26	0.2937	-0.53	0.0039
P	-0.05	0.1671	-0.05	0.0249
S	-0.01	0.5864	-0.01	0.1831
MICRO-NUTRIENTS (mg/kg)				
B	-4.56	0.2670	-5.68	0.0215
Cu	-2.32	0.1162	-2.31	0.0034
Fe	1.96	0.4980	0.14	0.9560
Mn	2.28	0.0647	2.17	0.0566
Mo	-2.15	0.0779	-3.06	0.0029
Zn	-2.66	0.3864	0.93	0.6247
SEED DISEASE (%)				
Cercospora	1.44	0.4262	-0.22	0.6727
Charcoal rot	0.41	0.4243	—	—
Fusarium	-6.94	0.0890	0.00	1.0000
Phomopsis	4.62	0.3104	0.09	0.9224

*Mean of genotypes with <80% germination minus the mean of genotypes with \geq 80% germination for each year. In 2012 there were three genotypes with \geq 80% germination and in 2013 there were six genotypes \geq 80% germination (see **Table 1**).

The experiment was conducted in 2012 and 2013 at the Jamie Whitten Delta States Research Center, Stoneville, MS.

Illinois exotic pool were derived from PIs 68508, 445837, 361064, 407710, 189930, and 68600 and previously selected for high yield potential, and the four breeding lines from the Mississippi exotic pool were derived from PIs 587982A and 603756 and previously selected for high germinability under heat stress. Given the distinct nature of the three germplasm pools, some inter-pool comparisons are in order. First, there was no significant difference between the three pools for seed yield or total seed damage in either year (**Table 11**). The high germinability (MS) pool had significantly lower seed size than the other two pools in 2013, but not in 2012 (**Table 11**). The MS pool had lower hard seed and higher germination and AA than the cultivar pool in both years (**Table 11**). There was less seed wrinkling in the MS pool than in the cultivar pool only in 2013, but the MS pool had lower green seed damage than the other two pools in both years (**Table 11**). In terms of seed constituents, the MS pool had higher protein than the other two pools in both years, while having lower Ca than the cultivar pool in both years (**Table 11**). The MS pool had higher oleic acid than the other two pools only in 2012 (**Table 11**). Given the increasing world-wide demand for high quality protein meal, and given the increased likelihood of global warming, the differences between pools for protein content are striking (**Tables 4, 11**). During the breeding process, the four lines in the high germinability MS pool were never selected for protein content. Rather, they were repeatedly selected in a pedigree breeding protocol that tested only seed germination from the F₂ through F₅ generations. Yet, as can be observed in **Table 4**, their protein contents are significantly and substantially higher than any other line in either of the other germplasm pools. That all four lines have high protein content is likely due to more than just chance. The potential association between protein content and high germinability needs further investigation.

A trait of high interest to producers in a dryland system is the maturity that will maximum yield. A later maturity may have the potential to utilize more sunlight and maximize more of the growing season, but an earlier maturity might better utilize early season rains and avoid late season droughts. The current research provides two different non-irrigated years (environments) to suggest an answer. It is interesting that neither the earliest (MG II, Dwight) nor latest (MG V, AG5606) maturity extreme gave the best return for yield (**Table 3**). In both years, the highest yielding lines were late IIIs to mid-to-late IVs. In 2012, for the 13 April planting date, the three highest yielding lines matured from 13 August (AG3905, 3797 kg/ha) to 14 August (LG04-1459-6, 3775 kg/ha) to 26 August (DK4866, 4122 kg/ha; **Tables 1, 3**). For the 30 April planting date in 2013, the three highest yielding lines matured from 24 August (AG3803, 2865 kg/ha and AG3905, 2670 kg/ha) to 7 September (DK4866, 2428 kg/ha; **Tables 1, 3**). Late MG III AG3905 and mid MG IV DK4866 were included in the highest yielding three lines in each year. In the most stressful year (2013), the late IIIs (AG3803 and AG3905) were the highest yielding. However, in the less stressful year (2012), the IVs (DK4866 and LG04-1459-6) were the highest yielding (**Table 3**). It might therefore make sense for producers utilizing a dryland

TABLE 11 | Inter-pool comparisons in the three germplasm pools [(cultivars, Mississippi (MS) exotic breeding lines, Illinois (IL) exotic breeding lines)].

Year	2012	2012	2012	2013	2013	2013
Group	Cultivar	IL breeding*lines	MS breeding lines	Cultivar	IL breeding lines	MS breeding lines
Kg/ha	3266a	3112a	2650a	2183a	2218a	1939a
g/100 seed	11.3a	12.4a	10.1a	9.8a	10.1a	8.2b
Flowering date	53a	50a	68b	71a	67a	87b
Full maturity (R8)	144ab	129a	155a	155a	144a	161a
Harvesting date	145ab	131a	158a	158a	146a	163a
Hard seed	20.9a	7.8b	0.8b	26.3a	9.3b	0.8b
Germination	47.6b	68.4a	76.8a	56.1b	78.2a	90.8a
Accelerated aging	43.3b	45.0ab	67.0a	59.3b	74.9a	77.2a
Total seed damage	0.54a	0.36a	0.48a	0.37a	0.81a	0.28a
Wrinkle	28.5a	21.1a	18.3a	23.3a	23.3ab	6.7b
Green Seed	18.9a	25.6a	0.0b	18.9a	25.6a	2.5b
Protein	38.0b	35.9c	42.9a	37.2a	35.1b	40.7c
Oil	21.9b	23.8a	20.4c	21.3b	22.7a	20.8b
Palmitic	10.2a	10.1a	10.3a	11.0b	10.7b	11.7a
Stearic	4.2a	4.2a	4.1a	4.5a	4.4a	4.5a
Oleic	27.9b	27.1b	32.1a	24.2a	24.2a	27.2a
Linoleic	53.9a	54.4a	50.6b	56.4a	56.6a	53.7b
Linolenic	5.1a	5.8a	4.9a	6.2a	6.6a	5.1b
Ca	0.43a	0.34b	0.24c	0.45a	0.40ab	0.31b
K	1.70b	1.84b	2.29a	1.91a	2.05b	2.14b
Mg	0.25a	0.25a	0.22b	0.26a	0.25a	0.25a
P	0.49b	0.49b	0.60a	0.55b	0.53b	0.64a
C	49.50b	50.54a	48.87c	52.94a	53.50b	52.74a
N	5.64b	5.34c	6.33a	5.74a	5.75a	6.51b
S	0.25b	0.25b	0.27a	0.31b	0.30b	0.33a
B	30.30b	29.82b	42.10a	33.04b	33.98b	42.07a
Cu	14.76b	14.71b	19.46a	16.87b	16.44b	20.16a
Fe	67.17a	64.60a	67.32a	82.51a	82.09a	85.22a
Mn	21.20a	19.29a	19.83a	23.95a	20.80b	21.72ab
Mo	4.08b	3.56b	7.68a	5.06b	4.76b	9.77a
Zn	48.39b	44.91b	55.60a	60.54a	56.51a	61.60a
Mold	1.10a	0.00a	0.00a	3.00a	1.10a	0.80a
Purple seed	0.70a	0.00a	0.00a	0.00a	0.00a	0.00a
Phomopsis	8.90a	8.00a	15.70a	1.20a	2.20a	0.30a
Cercospora	3.40a	0.00a	2.30a	0.70a	1.30a	0.30a
Fusarium	1.80a	0.00a	9.00a	0.70a	0.40a	0.70a
<i>Macrophomena phaseolina</i>	0.10b	1.30a	0.00b	0.00b	0.00b	0.00b

*MS breeding lines, Mississippi breeding lines; IL breeding lines, Illinois breeding lines; R8, full maturity stage; within a year, trait, and row, means with the same letter are not significantly different.

The experiment was conducted in 2012 and 2013 at the Jamie Whitten Delta States Research Center, Stoneville, MS.

Means within a row followed by the same letter are not significantly different at the 5% level as determined by Fishers' LSD test.

production system to plant a strategic mix of both late IIIs and IVs.

CONCLUSIONS

This research demonstrated that two genotypes (25-1-1-4-1-1 and 34-3-1-2-4-1) with $\geq 80\%$ germinability showed higher seed protein content, although there was no correlation shown

between germination and protein across all genotypes. Compared with the checks, seed of two genotypes with $\geq 80\%$ germinability (25-1-1-4-1-1 and 34-3-1-2-4-1) maintained significantly higher levels of N, P, B, Cu, and Mo, reflecting the possible roles of these nutrients for seed germination and their overall beneficial effects on seed health. The line 25-1-1-4-1-1 will be released as germplasm in 2017 and planned to be given to a public entity by Material Transfer Agreement (MTA); the line 34-3-1-2-4-1 was given to an industry entity by MTA and will be considered

for release in the future. These two lines represent the first stage of incorporating high germination traits into cultivars for soybean producers. The correlation between germination and some minerals such as B and Mo in 1 year only may reflect the different responses of nutrients to the growing environment, in our case, drought and high temperature. Seed germinability could partially be affected by hard seed and Ca level, as high germinability genotypes showed lower hard seed and lower Ca content compared to other genotypes. The general low levels of seed infection in most of the genotypes may indicate that these genotypes have some tolerance to these diseases, but more likely it is indicative of dryland growing conditions, where non-irrigated production systems are likely to have fewer seed diseases. In the water stress year (2013), a high germinability genotype (4025-1-1-4-1-1) showed moderately high yield (the fourth highest yielding genotype that year), which may indicate that it has drought stress tolerance. Further, research is needed to select for both high germination and high yield under drought and heat stress. This research will be beneficial to soybean breeders selecting for soybean seed with high seed nutritional values and high germination under dryland conditions.

REFERENCES

- Adams, C. A., Fjerstad, M. C., and Rinne, R. W. (1983). Characteristics of soybean maturation: necessity for slow dehydration. *Crop Sci.* 23, 265–267. doi: 10.2135/cropsci1983.0011183X002300020021x
- Andrade, E. R., Ribeiro, V. N., Azevedo, C. V. G., Chiorato, A. F., Williams, T. C. R., and Carbonell, S. A. M. (2016). Biochemical indicators of drought tolerance in the common bean (*Phaseolus vulgaris* L.). *Euphytica* 210, 277–289. doi: 10.1007/s10681-016-1720-4
- AOAC (Association of Official Analytical Chemists) (1990a). “Method 988.05,” in *Official Methods of Analysis, 15th Edn.*, ed K. Helrich (Arlington, VA: Elsevier Science Publishing), 70.
- AOAC (Association of Official Analytical Chemists) (1990b). “Method 920.39,” in *Official Methods of Analysis, 15th Edn.*, ed K. Helrich (Arlington, VA: Elsevier Science Publishing), 71.
- Association of Official Seed Analysts (2001). *Rules for Testing Seeds*. Las Cruces, NM.
- Association of Official Seed Analysts (2002). *Seed Vigor Testing Handbook*. Las Cruces, NM.
- Bandemer, S. L., and Schaible, P. J. (1944). Determination of iron. A study of the ophenanthroline method. *Ind. Eng. Chem. Anal. Ed.* 16, 317–319. doi: 10.1021/i560129a013
- Barnett, H. L., and Hunter, B. B. (1998). *Illustrated Genera of Imperfect Fungi*. St. Paul, MN: American Phytopathological Society.
- Bellaloui, N., Bruns, H. A., Abbas, H. K., Mengistu, A., Fisher, D. K., and Reddy, K. N. (2015). Effects of row-type, row-spacing, seeding rate, soil-type, and cultivar differences on soybean seed nutrition under US Mississippi delta conditions. *PLoS ONE* 10:e0129913. doi: 10.1371/journal.pone.0129913
- Bellaloui, N., Hanks, J. E., Fisher, D. K., and Mengistu, A. (2009b). Soybean seed composition is influenced by within-field variability in soil nutrients. *Plant Manag. Netw. Online Crop Manag.* doi: 10.1094/CM-2009-1203-01-RS. (Accessed April 2, 2017).
- Bellaloui, N., Mengistu, A., and Paris, D. R. (2008). Soybean seed composition in cultivars differing in resistance to charcoal rot (*Macrophomina phaseolina*). *J. Agric. Sci.* 146, 667–675. doi: 10.1017/S0021859608007971
- Bellaloui, N., Mengistu, A., Fisher, D. K., and Abel, C. A. (2012a). Soybean seed composition constituents as affected by drought and phomopsis in

AUTHOR CONTRIBUTIONS

NB contributed to the planning, design, analysis, interpretation, and writing. JS contributed to the planning, design, data interpretation, and writing. AM contributed to the analysis, data interpretation, and writing. JR contributed to the planning, design, analysis, data interpretation, and writing. AG contributed to the analysis, data interpretation, writing, revising the manuscript critically with intellectual content.

ACKNOWLEDGMENTS

We thank Sandra Mosley for lab analysis, and Philip Handly and Hans Hinrichsen for field management. This work was funded in part by USB under project number 1420-532-5650 under the title “Increasing Soybean Yield with Exotic Germplasm,” and also funded by the U.S. Department of Agriculture, Agricultural Research Service Project 6402-21220-012-00D. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture.

- phomopsis susceptible and resistant genotypes. *J. Crop. Improv.* 26, 428–453. doi: 10.1080/15427528.2011.651774
- Bellaloui, N., Mengistu, A., Walker, E. R., and Young, L. D. (2014). Soybean seed composition as affected by seeding rates and row spacing. *Crop Sci.* 54, 1782–1795. doi: 10.2135/cropsci2013.07.0463
- Bellaloui, N., Smith, J. R., Gillen, A. M., and Ray, J. D. (2010). Effect of maturity on seed sugars as measured on near-isogenic soybean (*Glycine max*) lines. *Crop Sci.* 50, 1978–1987. doi: 10.2135/cropsci2009.10.0596
- Bellaloui, N., Smith, J. R., Gillen, A. M., and Ray, J. D. (2011). Effects of maturity, genotypic background, and temperature on seed mineral composition in near-isogenic soybean lines in the early soybean production system. *Crop Sci.* 51, 1161–1171. doi: 10.2135/cropsci2010.04.0187
- Bellaloui, N., Smith, J. R., Gillen, A. M., Fisher, D. K., and Mengistu, A. (2012b). Effect of shade on seed protein, oil, fatty acids, and minerals in soybean lines varying in seed germinability in the Early Soybean Production System. *Am. J. Plant Sci.* 3, 84–95. doi: 10.4236/ajps.2012.31008
- Bellaloui, N., Smith, J. R., Ray, J. D., and Gillen, A. M. (2009a). Effect of maturity on seed composition in the early soybean production system as measured on near-isogenic soybean lines. *Crop Sci.* 49, 608–620. doi: 10.2135/cropsci2008.04.0192
- Bishnoi, U. R., Kaur, G., and Khan, M. H. (2007). Calcium, phosphorus, and harvest stages effect soybean seed production and quality. *J. Plant Nutr.* 30, 2119–2127. doi: 10.1080/01904160701700608
- Boydak, E., Alpaslan, M., Hayta, M., Gerçek, S., and Simsek, M. (2002). Seed composition of soybeans grown in the Harran region of Turkey as affected by row spacing and irrigation. *J. Agric. Food Chem.* 50, 4718–4720. doi: 10.1021/jf0255331
- Bradley, C. A., Hartman, G. L., Wax, L. M., and Pedersen, L. W. (2002). Quality of harvested seed associated with soybean cultivars and herbicides under weed-free conditions. *Plant Dis.* 86, 1036–1042. doi: 10.1094/PDIS.2002.86.9.1036
- Cavell, A. J. (1955). The colorimetric determination of phosphorus in plant materials. *J. Sci. Food Agric.* 6, 479–480. doi: 10.1002/jsfa.2740060814
- Chebroli, K. K., Fritsch, F. B., Ye, S., Krishnan, H. B., Smith, J. R., and Gillman, J. D. (2016). Impact of heat stress during seed development on soybean seed Metabolome. *Metabolomics* 12, 1–14. doi: 10.1007/s11306-015-0941-1
- Chen, P., Huhn, M. R., Buss, G. R., Gunduz, I., and Wilson, J. H. (2001). “Interrelationship among seed quality traits of specialty soybeans for natto and tofu,” in *Paper Presented at: ASA-CSSA-SSSA International Annual Meetings* (Charlotte, NC).

- Childs, C. E., and Henner, E. B. (1970). A Direct comparison of the Pregl, Dumas, Perkin-Elmer, and Hewlett-Packard(F&M) carbon-hydrogen-nitrogen procedures. *Microchem. J.* 15, 590–597. doi: 10.1016/0026-265X(70)90102-5
- Diers, B. W., Cary, T. R., Thomas, D. J., and Nickell, C. D. (2006). Registration of 'LD00-3309' soybean. *Crop Sci.* 46, 1384. doi: 10.2135/cropsci2005.06.0164
- Dordas, C., Apostolides, G., and Goundra, O. (2007). Boron application affects seed yield and seed quality of sugarbeets. *J. Agric. Sci.* 145, 377–384. doi: 10.1017/S0021859607006879
- Dornbos, D. L. Jr., and Mullen, R. E. (1992). Soybean seed protein and oil contents and fatty acid composition adjustments by drought and temperature. *JAOS* 69, 228–231. doi: 10.1007/bf02635891
- Dumas, J. B. A. (1831). Procédes de l'analyse organique. *Ann. Chim. Phys.* 247, 198–213.
- Durham, D. (2003). United soybean board. The United soybean board's better bean initiative: building united states soybean competitiveness from the inside out. *Ag. Bio. Forum* 6, 23–26. Available online at: <http://www.agbioforum.org/v6n12/v6n12a07-durham.htm>
- Federal Grain Inspection Service (FGIS) (2013). *Grain Inspection Handbook Book II: Soybean*. Available online at: <https://www.gipsa.usda.gov/fgis/handbook/grain-insp/grbook2/corn.pdf>
- Fehr, W. R., and Caviness, C. E. (1977). *Stages of Soybean Development*. Iowa Agric. Exp. Stn. Spec. Rep. 80. Iowa State University, Ames, IA.
- Franca-Neto, J. B., Krzyzanowski, F. C., Henning, A. A., West, S. H., and Miranda, L. C. (1993). Soybean seed quality as affected by shriveling due to drought stresses during seed filling. *Seed Sci. Technol.* 21, 107–116.
- Franca-Neto, J. B., West, S. H., and Vaughan, W. R. (1988). Multiple quality evaluation of soybean seed produced in Florida in 1986. *Soil Crop Sci. Soc. Fla. Proc.* 47, 201–206.
- Gibson, L. R., and Mullen, R. E. (1996). Soybean seed quality reductions by high day and night temperatures. *Crop Sci.* 36, 1615–1619. doi: 10.2135/cropsci1996.0011183X003600060034x
- Gillen, A. M., and Shelton, G. W. (2011). *Uniform Soybean Tests Southern States. Annual Reports*. Available online at: <https://www.ars.usda.gov/SP2UserFiles/Place/60661000/UniformSoybeanTests/2011SoyBook.pdf> (Accessed July 23, 2016).
- Grain inspection handbook Book II Soybean (2013). Available online at: <https://www.gipsa.usda.gov/fgis/handbook/grain-insp/grbook2/soybean.pdf> (Accessed August 3, 2016).
- Green, D. E., Pinell, E. L., Cavanah, L. E., and Williams, L. F. (1965). Effect of planting date and maturity on soybean seed quality. *Agron. J.* 57, 165–168. doi: 10.2134/agronj1965.00021962005700020010x
- Griehop, C. M., and Fahey, G. C. Jr. (2001). Comparison of quality characteristics of soybeans from Brazil, China, and the United States. *J. Agric. Food Chem.* 49, 2669–2673. doi: 10.1021/jf0014009
- Haq, M. U., and Mallarino, A. P. (2005). Response of soybean grain oil and protein concentrations to foliar and soil fertilization. *Agron. J.* 97, 910–918. doi: 10.2134/agronj2004.0215
- Heatherly, L. G. (1999). "Early soybean production system (ESPS)," in *Soybean Production in the Midsouth*, eds L. G. Heatherly and H. F. Hodges (Boca Raton, FL: CRC Press), 103–118.
- Hepperly, P. R., and Sinclair, J. B. (1978). Quality losses in phomopsis-infected soybean seeds. *Phytopathology* 68, 1684–1687. doi: 10.1094/Phyto-68-1684
- Hobbs, T. W., Schmitthenner, A. F., and Kuter, G. A. (1985). A new *Phomopsis* species from soybean. *Mycologia* 77, 534–544. doi: 10.2307/3793352
- Holmes, F. L. (1963). Elementary analysis and the origins of physiological chemistry. *Isis* 54, 50–81. doi: 10.1086/349664
- John, M. K., Chuah, H. H., and Neufeld, J. H. (1975). Application of improved azomethine-H method to the determination of boron in soils and plants. *Anal. Lett.* 8, 559–568. doi: 10.1080/00032717508058240
- Karr-Lilienthal, L. K., Griehop, C. M., Merchen, N. R., Mahan, D. C., and Fahey, G. C. (2004). Chemical composition and protein quality comparisons of soybeans and soybean meals from five leading soybean-producing countries. *J. Agric. Food Chem.* 52, 6193–6199. doi: 10.1021/jf049795+
- Kebede, H., Smith, J. R., and Ray, J. D. (2013). A new gene that controls seed coat wrinkling in soybean. *Euphytica* 189, 309–320. doi: 10.1007/s10681-012-0818-6
- Kebede, H., Smith, J. R., and Ray, J. D. (2014). Identification of a single gene for seed coat impermeability in soybean PI 594619. *Theor. Appl. Genet.* 127, 1991–2003. doi: 10.1007/s00122-014-2355-2
- Keiser, J. R., and Mullen, R. E. (1993). Calcium and relative humidity effects on soybean seed nutrition and seed quality. *Crop Sci.* 33, 1345–1349. doi: 10.2135/cropsci1993.0011183X003300060044x
- Keith, B. C., and Delouche, J. C. (1999). "Seed quality, production, and treatment," in *Soybean Production in the Midsouth*, eds L. G. Heatherly and H. G. Hodges (Boca Raton, FL: CRC Press), 197–230.
- Kilen, T. C., and Hartwig, E. E. (1978). An inheritance study of impermeable seed in soybeans. *Field Crops Res.* 1, 65–70. doi: 10.1016/0378-4290(78)90007-2
- Kmetz, K., Ellet, C. W., and Schmitthenner, A. F. (1974). Isolation of seed-borne *Diaporthe phaseolorum* and *Phomopsis* from immature soybean plants. *Plant Dis. Rep.* 58, 978–982.
- Kmetz, K., Ellet, C. W., and Schmitthenner, A. F. (1979). Soybean seed decay: source of inoculum and nature of infection. *Phytopathology* 69, 798–801. doi: 10.1094/Phyto-69-798
- Kmetz, K., Schmitthenner, A. F., and Ellett, C. W. (1978). Soybean seed decay: prevalence of infection and symptom expression caused by *Phomopsis* sp., *Diaporthe phaseolorum* var. *sojae*, and *D. phaseolorum* var. *cuilivora*. *Phytopathology* 68, 836–840. doi: 10.1094/Phyto-68-836
- Kulik, M. M., and Sinclair, J. B. (1999). "Phomopsis seed decay," in *Compendium of Soybean Diseases, 4th Edn.*, eds G. L. Hartman, J. B. Sinclair, and J. C. Rupe (St. Paul, MN: American Phytopathological Society), 31–32.
- Lamoine, R., Camera, S. L., Atanassova, R., Allario, T., Pourtau, N., Bonnemain, J. L., et al. (2013). Source-to-sink transport of sugar and regulation by environmental factors. *Front. Plant Sci.* 4:272. doi: 10.3389/fpls.2013.00272
- Lee, J. H., Kwon, J. H., Kang, J. W., Park, C. H., Tae, B. S., and Kim, S. W. (2009). Biodiesel production from various oils under supercritical fluid conditions by *Candida antarctica* lipase B using a stepwise reaction method. *Appl. Biochem. Biotechnol.* 156, 24–34. doi: 10.1007/s12010-008-8488-5
- LeVan, N. A., Goggi, A. S., and Mullen, R. (2008). Improving the reproducibility of the soybean standard germination test. *Crop Sci.* 48, 1933–1940. doi: 10.2135/cropsci2007.06.0372
- Loeppert, R. L., and Inskeep, W. P. (1996). "Colorimetric determination of ferrous iron and ferric iron by the 1,10-phenanthroline method," in *Methods of Soil Analysis: Part 3, Chemical Methods*, ed J. M. Bigham, (Madison, WI: SSSA), 659–661.
- Lohse, G. (1982). Microanalytical azomethine-H method for boron determination in plant tissue. *Commun. Soil Sci. Plant Anal.* 13, 127–134. doi: 10.1080/00103628209367251
- Marschner, H. (1995). *Mineral Nutrition of Higher Plants*. London: Academic Press.
- Mengel, K., and Kirkby, E. A. (1982). *Principles of Plant Nutrition*. Worblaufen Bern: International Potash Institute.
- Mengistu, A., and Heatherly, L. G. (2006). Planting date, irrigation, maturity group, year, and environment effects on phomopsis longicolla, seed germination, and seed health rating of soybean in the early soybean production system of the midsouthern USA. *Crop Prot.* 25, 310–317. doi: 10.1016/j.cropro.2005.05.011
- Mengistu, A., Castlebury, L. A., Smith, J. R., Ray, J. D., and Bellaloui, N. (2009). Seasonal progress of *phomopsis longicolla* infection on soybean plant parts and its relationship to seed quality. *Plant Dis.* 93, 1009–1018. doi: 10.1094/PDIS-93-10-1009
- Mengistu, A., Castlebury, L. A., Smith, J. R., Rossman, A. Y., and Reddy, K. N. (2007). Identification and pathogenicity of *Phomopsis* isolates from weed hosts and their effect on soybean. *Can. J. Plant Pathol.* 29, 283–289. doi: 10.1080/07060660709507471
- Mengistu, A., Smith, J. R., Bellaloui, N., Paris, R. L., and Wrather, J. A. (2010). Irrigation and time of harvest effects on evaluation of selected soybean accessions against *phomopsis longicolla*. *Crop Sci.* 50, 2055–2064. doi: 10.2135/cropsci2009.11.0657
- MSUCares (2016). *Mississippi State University, Extension Service*. Available online at: <http://ext.msstate.edu/anr/drec/weather.cgi> (Accessed June 22)
- Mullin, J. W., and Xu, W. (2001). Study of soybean seed coat components and their relationship to water absorption. *J. Agric. Food Chem.* 49, 5331–5335. doi: 10.1021/jf010303s

- Nakasathien, S. I., Israel, D. W., Wilson, R. F., and Kwanyuen, P. (2000). Regulation of seed protein concentration by supra-optimal nitrogen supply. *Crop Sci.* 40, 1277–1284. doi: 10.2135/cropsci2000.4051277x
- Nickell, C. D., Noel, G. R., Cary, T. R., and Thomas, D. J. (1998). Registration of 'Dwight' soybean. *Crop Sci.* 38, 1398. doi: 10.2135/cropsci1998.0011183x003800050046x
- Oltmans-Deardorff, S. E., Fehr, W. R., and Shoemaker, R. C. (2013). Marker-assisted selection for elevated concentrations of the α' subunit of β -conglycinin and its influence on agronomic and seed traits of soybean. *Crop Sci.* 53, 1–8. doi: 10.2135/cropsci2012.03.0205
- Orazaly, M., Chen, P., Zhang, B., and Zeng, A. (2014). Quantitative trait loci mapping for seed calcium content of soybean. *Crop Sci.* 54, 500–506. doi: 10.2135/cropsci2013.01.0026
- Paris, R. L., Mengistu, A., Tyler, J. M., and Smith, J. R. (2006). Registration of soybean germplasm line DT97-4290 with moderate resistance to charcoal rot. *Crop Sci.* 46, 2324–2325. doi: 10.2135/cropsci2005.09.0297
- Piper, E. L., and Boote, K. J. (1999). Temperature and cultivar effects of soybean seed oil and protein concentration. *J. Am. Oil Chem. Soc.* 76, 1233–1241. doi: 10.1007/s11746-999-0099-y
- Ray, J. D., Fritschi, F. B., and Heatherly, L. G. (2006). Large applications of fertilizer N at planting affects seed protein and oil concentration and yield in the Early Soybean Production System. *Field Crops Res.* 99, 67–74. doi: 10.1016/j.fcr.2006.03.006
- Rotundo, J. L., and Westgate, M. E. (2009). Meta-analysis of environmental effects on soybean seed composition. *Field Crops Res.* 110, 147–156. doi: 10.1016/j.fcr.2008.07.012
- Rotundo, J. L., Borrás, L., de Bruin, J. D., and Pedersen, P. (2014). Soybean nitrogen uptake and utilization in Argentina and United States cultivars. *Crop Sci.* 54, 1153–1165. doi: 10.2135/cropsci2013.09.0618
- Roy, K. W., Keith, B. C., and Andrews, C. H. (1994). Resistance of hardseeded soybean lines to seed infection by *Phomopsis*, other fungi, and soybean mosaic virus. *Can. J. Plant Pathol.* 16, 122–128. doi: 10.1080/07060669409500769
- Saio, K. (1976). Soybean resistant to water absorption. *Cereal Food World* 21, 168–173.
- Saio, K., Arai, K., and Watanabe, T. (1973). Fine structure of soybean seed coat and its changes on cooking. *Cereal Sci. Today* 18, 197–201.
- Sale, P. W. G., and Campbell, L. C. (1980). Patterns of mineral nutrient accumulation in soybean seed. *Field Crops Res.* 3, 157–163. doi: 10.1016/0378-4290(80)90021-0
- Salmeron, M., Gbur, E. E., Bourland, F. M., Buehring, N. W., Earnest, L., Fritschi, F. B., et al. (2014). Soybean maturity group choices for early and late plantings in the Midsouth. *Agron. J.* 106, 1893–1901. doi: 10.2134/agronj14.0222
- Samarah, N. H., and Mullen, R. E. (2004). Effect of maturity stage on seed germination and vigor of common vetch (*Vicia sativa* L.). *Seed Technol.* 26, 27–37.
- SAS Institute (2002–2010). SAS. Cary, NC: SAS Institute.
- Shi, A., Chen, P., Zhang, B., and Hou, A. (2010). Genetic diversity and association analysis of protein and oil content in food-grade soybeans from Asia and the United States. *Plant Breed.* 129, 250–256. doi: 10.1111/j.1439-0523.2010.01766.x
- Smith, J. R., Mengistu, A., Nelson, R. L., and Paris, R. L. (2008). Identification of soybean accessions with high germinability in high-temperature environments. *Crop Sci.* 48, 2279–2288. doi: 10.2135/cropsci2008.01.0026
- Spears, J. F., Tekrony, D. M., and Egli, D. B. (1997). Temperature during seed filling and soybean seed germination and vigor. *Seed Sci. Technol.* 25, 233–244.
- TeKrony, D. M., Egli, D. B., Balles, J., Tomes, L., and Stuckey, R. E. (1980). Effect of harvest maturity on soybean seed quality and *Phomopsis* sp. seed infection. *Crop Sci.* 24, 189–193. doi: 10.2135/cropsci1984.0011183X002400010045x
- TeKrony, D. M., Egli, D. B., Balles, J., Tomes, L., and Stuckey, R. E. (1984). Effect of date of harvest maturity on soybean seed quality and *Phomopsis* sp. seed infection. *Crop Sci.* 24, 189–193. doi: 10.2135/cropsci1984.0011183X002400010045x
- Tekrony, D. M., Grabau, L. J., DeLacy, M., and Kane, M. (1996). Early planting of early-maturing soybean: effects on seed germination and *Phomopsis* infection. *Agron. J.* 88, 428–433. doi: 10.2134/agronj1996.00021962008800030011x
- Thakur, M., and Hurburgh, C. R. (2007). Quality of U.S. soybean meal compared to the quality of soybean meal from other origins. *J. Am. Oil Chem. Soc.* 84, 835–843. doi: 10.1007/s11746-007-1107-8
- Thomison, P. R., Kenworthy, W. J., and McIntosh, M. S. (1990). *Phomopsis* seed decay in soybean isolines differing in stem termination, time of flowering, and maturity. *Crop Sci.* 30, 183–188. doi: 10.2135/cropsci1990.0011183X003000010040x
- Webb, A. M. (1999). Cell-mediated crystallization of calcium oxalate in plants. *Plant Cell* 11, 751–761. doi: 10.1105/tpc.11.4.751
- Wilcox, J. R., and Shibles, R. M. (2001). Interrelationships among seed quality attributes in soybean. *Crop Sci.* 41, 11–14. doi: 10.2135/cropsci2001.41111x
- Wilson, R. F. (2004). "Seed composition," in *Soybeans: Improvement, Production, and Uses, 3rd Edn.*, eds H. Boerma and J. E. Specht (Madison, WI: American Society of Agronomy, Crop Science Society of America, and Soil Science Society of America), 621–668.
- Zhang, B., Chen, P., Shi, A., Hou, A., Ishibashi, T., and Wang, D. (2009). Putative quantitative trait loci associated with calcium content in soybean seed. *J. Hered.* 100, 263–269. doi: 10.1093/jhered/esn096
- Zhang, M. W., Peng, Z. M., and Du, Y. Q. (1996). Combining ability and stability analysis of the content of trace elements Fe, Zn and Mn in special rice grains. *Chinese J. Rice Sci.* 10, 201–206.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Bellaloui, Smith, Mengistu, Ray and Gillen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Overexpression of the Starch Phosphorylase-Like Gene (*PHO3*) in *Lotus japonicus* has a Profound Effect on the Growth of Plants and Reduction of Transitory Starch Accumulation

Shanshan Qin^{1,2}, Yuehui Tang^{1,2}, Yaping Chen¹, Pingzhi Wu¹, Meiru Li¹, Guojiang Wu¹ and Huawu Jiang^{1*}

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University
of Mexico, Mexico

Reviewed by:

Ling Yuan,
University of Kentucky, USA
Matthew Paul,
Rothamsted Research –
Biotechnology and Biological
Sciences Research Council, UK
Alexandre Tromas,
National Autonomous University
of Mexico, Mexico

*Correspondence:

Huawu Jiang
hwjiang@scbg.ac.cn

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 20 April 2016

Accepted: 16 August 2016

Published: 31 August 2016

Citation:

Qin S, Tang Y, Chen Y, Wu P, Li M,
Wu G and Jiang H (2016)
Overexpression of the Starch
Phosphorylase-Like Gene (*PHO3*)
in *Lotus japonicus* has a Profound
Effect on the Growth of Plants
and Reduction of Transitory Starch
Accumulation.
Front. Plant Sci. 7:1315.
doi: 10.3389/fpls.2016.01315

¹ Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China, ² University of Chinese Academy of Sciences, Beijing, China

Two isoforms of starch phosphorylase (PHO; EC 2.4.1.1), plastidic PHO1 and cytosolic PHO2, have been found in all plants studied to date. Another starch phosphorylase-like gene, *PHO3*, which is an ortholog of *Chlamydomonas PHOB*, has been detected in some plant lineages. In this study, we identified three PHO isoform (*LjPHO*) genes in the *Lotus japonicus* genome. Expression of the *LjPHO3* gene was observed in all tissues tested in *L. japonicus*, and the *LjPHO3* protein was located in the chloroplast. Overexpression of *LjPHO3* in *L. japonicus* resulted in a drastic decline in starch granule sizes and starch content in leaves. The *LjPHO3* overexpression transgenic seedlings were smaller, and showed decreased pollen fertility and seed set rate. Our results suggest that *LjPHO3* may participate in transitory starch metabolism in *L. japonicus* leaves, but its catalytic properties remain to be studied.

Keywords: starch phosphorylase, gene expression, starch metabolism, pollen fertility, *Lotus japonicus* L.

INTRODUCTION

Starch phosphorylase (α -glucan phosphorylase, PHO; EC 2.4.1.1) catalyzes the reversible transfer of glucosyl units from glucose-1-phosphate to the non-reducing ends of α -1,4-D-glucan chains with the release of phosphate. Two major forms of PHO, the plastidic PHO1 or PHOL (which has a low affinity for glycogen) and the cytosolic PHO2 or PHOH (high glycogen affinity), have been observed across all the higher plants (Wirtz et al., 1980; Kruger and ap Rees, 1983). PHO1 has an additional 78–80 amino acid region (L78 domain) near the middle of the GT1_Glycogen_Phosphorylase domain. The L78 domain in PHO1 has a PEST region which serves as a signal for degradation in sweet potato (Chen et al., 2002; Lin et al., 2012). Mori et al. (1993) suggested that L78 domain in potato PHO1 lowered the affinity of the enzyme for large, branched substrates. Removal of the L70/L80 domain in rice PHO1 did not significantly alter the catalytic and regulatory properties of PHO1 but did affect heat stability (Hwang et al., 2016).

In *Arabidopsis*, mutants lacking PHS1 (plastidic PHO1) have normal patterns of diurnal starch metabolism and no significant changes in starch structure, indicating that this enzyme is not

essential for starch synthesis (SS) or degradation. However, the plants display increased sensitivity to drought stress and there is local accumulation of starch around stress-induced lesions (Zeeman et al., 2004). Genetic studies support a role for the PHS1 protein in transitory starch degradation (Malinova et al., 2014). In rice, the loss of PHO1 leads to a reduction in endosperm starch content and shrunken seeds when plants are grown at 20°C, though not when they are grown at 30°C (Sato et al., 2008). The cytosolic starch phosphorylase PHO2 may be involved, together with the cytosolic transglucosidase DPE2, in the metabolism of cytosolic maltose and heteroglycans formed by starch degradation in leaves (Fettke et al., 2006; Lu et al., 2006). In potato, altered levels of PHO2 mainly affect the molecular properties of the heteroglycans (Fettke et al., 2005). In *Chlamydomonas reinhardtii*, mutation in *STA4*, which encodes the PHOB protein, results in a significant reduction in starch content, the formation of abnormally shaped starch granules containing chain-length modified amylopectin and increased relative amounts of amylase under N-deficiency conditions (Dauvillée et al., 2006). These changes in starch content and structure indicate that PHOB plays a significant role during storage SS.

The ever-growing databases of genomic DNA sequences from “model” and “non-model” plants offer greatly enhanced opportunities to detect previously unknown gene families in higher plants. After comparing gene families involved in starch biosynthetic pathways in *Jatropha curcas* L. with those in *Arabidopsis* and other plants, we found that *J. curcas* L. and some dicots have an ACT domain-containing starch phosphorylase isoform (PHOA or PHO3) which had not previously been reported in higher plants. Phylogenetic analysis suggested that the putative PHO3 proteins form a new subclade with PHOB proteins from the green algae *Ostreococcus lucimarinus* and *C. reinhardtii* (Wu et al., 2015). In the study presented here we explored the function of the *PHO3* gene in *Lotus japonicus*. Our results revealed that the PHO3 protein was located in the chloroplast, and overexpression of the *PHO3* gene decreased starch accumulation in leaves and influenced plant growth and fertility in *L. japonicus*. The results presented here represent the first data on the function of the PHO3 protein subfamily in higher plants.

MATERIALS AND METHODS

Plant Growth and Bacterial Strains

Lotus japonicus genotype ‘MG-20’ was used as the wild-type control for phenotypic and genotypic analysis. Seeds were scarified for 10 min in sulfuric acid and planted in vermiculite irrigated with Broughton and Dilworth (B&D) nutrient solution without nitrogen. After planting, seedlings were inoculated with *Mesorhizobium loti* MAFF303099. For testing, all plants were grown in a growth chamber (day/night cycles of 18 h/6 h; temperature 22°C/20°C). For harvesting seeds, seedlings were planted in peat-vermiculite inoculated with *M. loti* MAFF303099 and irrigated with B&D nutrient solution without nitrogen.

Sequence Retrieval and Analysis

Sequences of PHO proteins were retrieved from GenBank¹. *Arabidopsis* PHO proteins were used as queries in BLAST searches against the *L. japonicus* genome database². Conserved motifs in PHO proteins were analyzed using the NCBI's CDD database³ (Marchler-Bauer et al., 2015). The plastid transit peptide cleavage site of the PHO proteins was predicted using the program TargetP Server v1.01⁴ (Emanuelsson et al., 1999). For phylogenetic analysis, multiple sequence alignments of PHO amino acid sequences were performed using ClustalW. The tree was constructed using the neighbor-joining (NJ) method and 100 bootstraps in order to group putative full-length PHO amino acid sequences, and the results were displayed with Mega software version 4 (Tamura et al., 2007).

Plasmid Constructs and Plant Transformation

For construction of the overexpression vector, full length *LjPHOA* cDNA was amplified by RT-PCR using the primers given in Supplementary Table S1. After digestion by restriction enzymes, the cDNA fragments were cloned into the *Kpn* I/*Xba* I sites of pCambia 1302 behind the 35S promoter. The resulting construct was introduced into *Agrobacterium tumefaciens* strain AGL1 by the freeze-thaw procedure. Transformation of *L. japonicus* was carried out according to the method described by Chen et al. (2014).

For subcellular location analysis, the complete coding sequences with the exception of the stop codon were amplified by PCR with the primers given in Supplementary Table S1. After digested by restriction enzymes, the cDNA fragments were cloned into the *Bam* HI/*Xho* I sites of pSAT-EYFP-N1 upstream of the *EYFP* gene. Protoplasts of *Arabidopsis* were isolated and transformed according to the method described by Yoo et al. (2007).

For protein expression in *Escherichia coli*, three fragments, P1 (Δ^{41} PHO), P2 (Δ^{183} PHO), and P3 (complete CDS) were amplified by PCR with the primers given in Supplementary Table S1. The PCR products were digested and introduced into the pGEX-KG vector at the sites of *Xba* I/*Xho* I (P1 and P3) or *Nco* I/*Xho* I (P2) on the 3'-terminus of the GST gene (without stop codon).

Expression of *LjPHO* in *E. coli*

Cultures of *E. coli* strain Rosetta containing pGEX-KG (native plasmid), P1, P2, and P3 were grown in Luria-Bertani medium. Overnight cultures were inoculated into fresh medium at a 1:100 dilution and grown at 37°C until the A600 was 0.6. IPTG was added to 0.5 mM and the cultures were grown for 8 h at 22°C. Cells were collected from 400 mL cultures by centrifugation, suspended in one-twentieth culture volume of sonication buffer (50 mM Tris-acetate, pH 7.5, 10 mM EDTA, and 5 mM DTT), and broken by sonication. Lysates were cleared by centrifugation

¹<http://www.ncbi.nlm.nih.gov/>

²<http://www.kazusa.or.jp/lotus/index.html>

³<http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>

⁴<http://www.cbs.dtu.dk/services/TargetP/>

at $10,000 \times g$ for 10 min, and the supernatants (crude enzyme extracts) were used for subsequent analyses. The proteins were separated by 7.5% SDS-PAGE. Gels were stained with Coomassie Brilliant Blue R-250 (Jiang et al., 2003).

The recombinant proteins were purified using the GST-BIND™ Resin (Novagen®, cat. No. 70541)⁵ following the manufacturer's instructions. GST-LjPHO3 recombinant protein fractions were selected after SDS-PAGE analysis and pooled together. The proteins were precipitated by slowly added two volumes of saturated ammonium sulfate solution pre-chilled at 4°C. After precipitation by centrifugation, the pellet was resuspended with 1 ml of 25 mM HEPES-NaOH buffer (pH 7.0; containing 10% glycerol). The protein solution dialyzed against 3×400 mL of the 25 mM HEPES-NaOH buffer at 4°C for 24 h. After clarification by centrifugation the pure enzyme preparation was stored at -80°C until used for analysis. For enzyme assay, both crude enzyme extracts and purified recombinant proteins were tested, respectively (Zeeman et al., 2004). The phosphorylase b from rabbit muscle (P6635, sigma) was used as a positive control to ensure the reaction system for the activity determination was adopted.

RNA Isolation and qRT-PCR

Samples used for expression analysis were: leaves, roots, stems, and nodules from 3-week-old seedlings; whole unexpanded flowers; young siliques 1–1.5 cm in length; and developing seeds 15–20 days after flowering. Total RNA was extracted from tissues of *L. japonicus* using an RNeasy Plant Mini Kit (QIAGEN)⁶ following the manufacturer's instructions, and the isolated RNA was treated with RNase-free DNase I (Roche)⁷. First-strand cDNA was synthesized from 2 µg RNA using M-MLV reverse transcriptase (Promega)⁸ according to the manufacturer's instructions. Primer pairs for the PHO genes were designed by the Primer3 software⁹. Pairs of primers are selected which can give specific DNA amplification by PCR. *Ubiquitin* (GenBank accession No. AFK37806) was used as a reference gene. The primers are given in Supplementary Table S1. QRT-PCR was performed on a Mini Option real-time PCR system (LightCycler 480). Cycling conditions were as follows: 95°C for 30 s, 95°C for 5 s, 60°C for 20 s, and 72°C for 20 s. The reaction was performed for 40 cycles. The experiment was performed with three biological replicates and average values are presented (Chen et al., 2014).

Laser Scanning Confocal Microscopy and Electron Microscopy

Fluorescence images were recorded with a laser scanning confocal microscope (LSM510 META, Zeiss¹⁰). eYFP fluorescence was imaged at an excitation wavelength of 514 nm (30% power), and the emission wavelength of 527 nm.

Chlorophyll fluorescence was imaged at an excitation wavelength of 543 nm, and the emission wavelength of 562 nm.

To obtain ultra-thin sections, leaf four or five from the top of the main branch in 10-week seedlings was fixed in 2.5% glutaraldehyde and 2% paraformaldehyde, then dehydrated in an ethanol series and embedded in Spurr resin. Ultra-thin sections (0.1 µm) were stained with 2% uranyl acetate for 1 h and 6% lead citrate for 20 min and observed with an electron microscope (JEM-1010, Jeol¹¹; Jiang et al., 2007).

Analysis of Starch and Soluble Sugars

Leaves (100 mg fresh weight, hand-homogenized using liquid nitrogen) were extracted three times with 1 ml of ethanol (80% v/v) for 10 min at 80°C. The supernatant after each extraction was recovered by centrifugation. After the last extraction, the pellet was washed with 0.5 ml of 80% ethanol. All supernatants were transferred to a test tube and the combined volume was adjusted to 5 ml with 80% ethanol. Soluble sugar content was determined by a colorimetric method with a sulfuric acid-phenol reagent using a sucrose standard curve (DuBois et al., 1951). The remaining ethanol-insoluble residue was extracted twice by suspension in 1 ml purified water and incubation at 100°C for 30 min; the supernatants were recovered by centrifugation. The supernatants were transferred to a test tube and the combined volume was adjusted to 5 ml with purified water. The starch content was measured by a colorimetric method using an iodine solution and calculated according to a potato starch standard curve.

Preparation of Enzymes from Leaves and Enzyme Measurements

All procedures were performed at 0–4°C. Fresh leaves (100 mg) were hand-homogenized with a glass homogenizer in 500 µL of solution containing 100 mM HEPES-NaOH (pH 7.5), 1 mM EDTA, 5 mM DTT, 10 mM MgCl₂, 20 mM KCl, and 10% (v/v) glycerol. The homogenate was centrifuged at $12,000 \times g$ for 10 min, and the pellet was washed ($250 \mu\text{L} \times 2$) with the same buffer. The resulting supernatants were used for the preparation of enzymes.

The ADP-glucose pyrophosphorylase (AGPase) assay was carried out according to the method described by Nishi et al. (2001). The SS and branching enzyme (BE) assay was according to the method described by Jiang et al. (2003). The starch phosphorolysis activity of starch phosphorylase was assayed using a continuous assay in the direction of Glc-1-P formation, coupled to the production of NADH (Zeeman et al., 2004). The SS activity of starch phosphorylase was detected on native glycogen-containing zymograms (Dauvillée et al., 2006).

Observation of Pollen Morphology and Germination

For the observation of pollen morphology and fertility, pollen grains were harvested from fully expanded flowers. After staining with iodine-potassium iodide solution, pollen was viewed with

⁵<http://www.merckmillipore.com/>

⁶<http://www.qiagen.com>

⁷<http://www.roche.com>

⁸<http://au.promega.com>

⁹<http://primer3.ut.ee/>

¹⁰<http://www.zeiss.com/>

¹¹<http://www.jeol.com>

a bright-field microscope. To test the pollen germination rate, pollen grains were dispersed in a solution of 10% sucrose, 0.01% H_3BO_3 , 0.05% $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$. After incubating at 37°C for 30 min, germinated pollen grains were viewed with a bright-field microscope.

RESULTS

Identification and Characterization of the Putative Starch Phosphorylase Genes in the *L. japonicus* Genome

TBLASTN searches of the *L. japonicus* genome¹² using the amino acid sequences of *Arabidopsis* PHO proteins revealed four putative *PHO* genes. The *L. japonicus* *PHO* genes were designated *LjPHO1;1* (Lj2g3v1079510), *LjPHO1;2* (Lj0g3v0360239),

LjPHO2 (LjB08M07.90.r, database build 2.5), and *LjPHO3* (Lj6g3v2006830), respectively, on the basis of the unrooted NJ-tree (Supplementary Figure S1) constructed from PHO proteins of selected plants. Alignment of the genomic DNA sequences with the cDNA sequences reveals that the *LjPHO* genes contain 15–20 exons separated by 14–19 introns within their coding domain sequences (Supplementary Table S2). *LjPHO2* contains two additional introns in its 5' untranslated region. All *LjPHO* proteins contain the GT1_Glycogen_Phosphorylase domain (cd04300). *LjPHO1* and *LjPHO3*, but not *LjPHO2*, contain a putative plastid-targeting peptide region (TP) as predicted by ChloroP 1.1¹³ (Figure 1A). *LjPHO3* contains an ACT domain (cl09141) in the N-terminal region. The two *LjPHO1* proteins have the L78 region near the middle of the GT1 domain; this region does not exist in the *LjPHO2* and *LjPHO3* proteins (Supplementary Figure S2).

¹²<http://www.kazusa.or.jp/lotus/blast.html>

¹³www.cbs.dtu.dk/services/ChloroP/

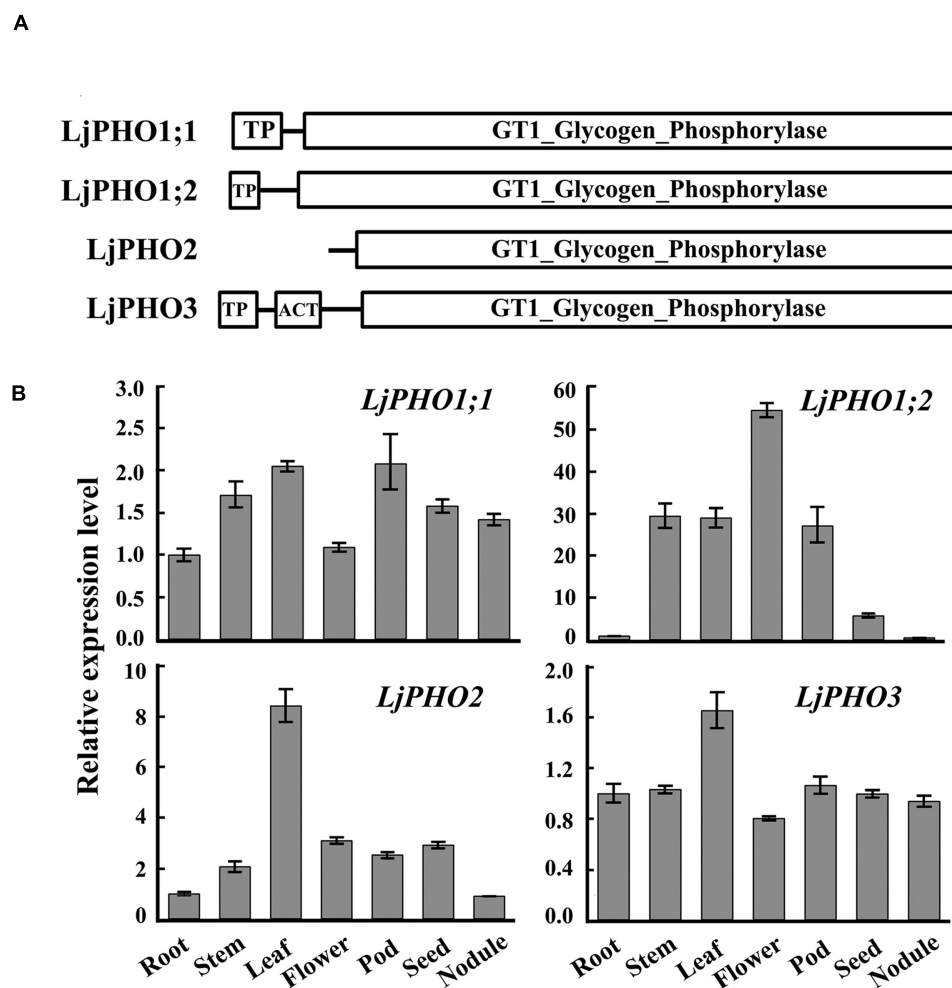


FIGURE 1 | PHO genes detected in the *Lotus japonicus* genome. (A) Comparison of the domains found within the predicted amino acid sequences of PHO proteins; **(B)** Expression analysis of the putative *PHO* genes. QRT-PCR of total RNA samples was used to measure putative *PHO* gene expression. Relative expression was normalized to the reference gene *ubiquitin* (internal control). Bars show means \pm SD of three biological replicates.

The expression levels of transcripts encoded by the *LjPHO* genes were measured by qRT-PCR in only one developmental stage, including leaf, root, stem, and nodule of 3-week-old plants, and flower, pod and seed of 10-week-old plants of *L. japonicus* MG-20. The results showed that the *LjPHO2* gene was expressed at the highest level in leaves, while *LjPHO1;2* was expressed weakly in roots, seeds, and nodules. The expression levels of *LjPHO1;1* and *LjPHO3* differed little among the tissues tested (Figure 1B).

The biochemical characteristics and biological functions of PHO1 and PHO2 proteins have been investigated in many plants. The present study focused on analysis of the function of the *LjPHO3* gene in *L. japonicus*.

Plastid Localization and Expression of *LjPHO3* in *E. coli*

To confirm the plastid localization of *LjPHO3*, as predicted by ChloroP 1.1, we examined the localization of *LjPHO3*-eYFP by

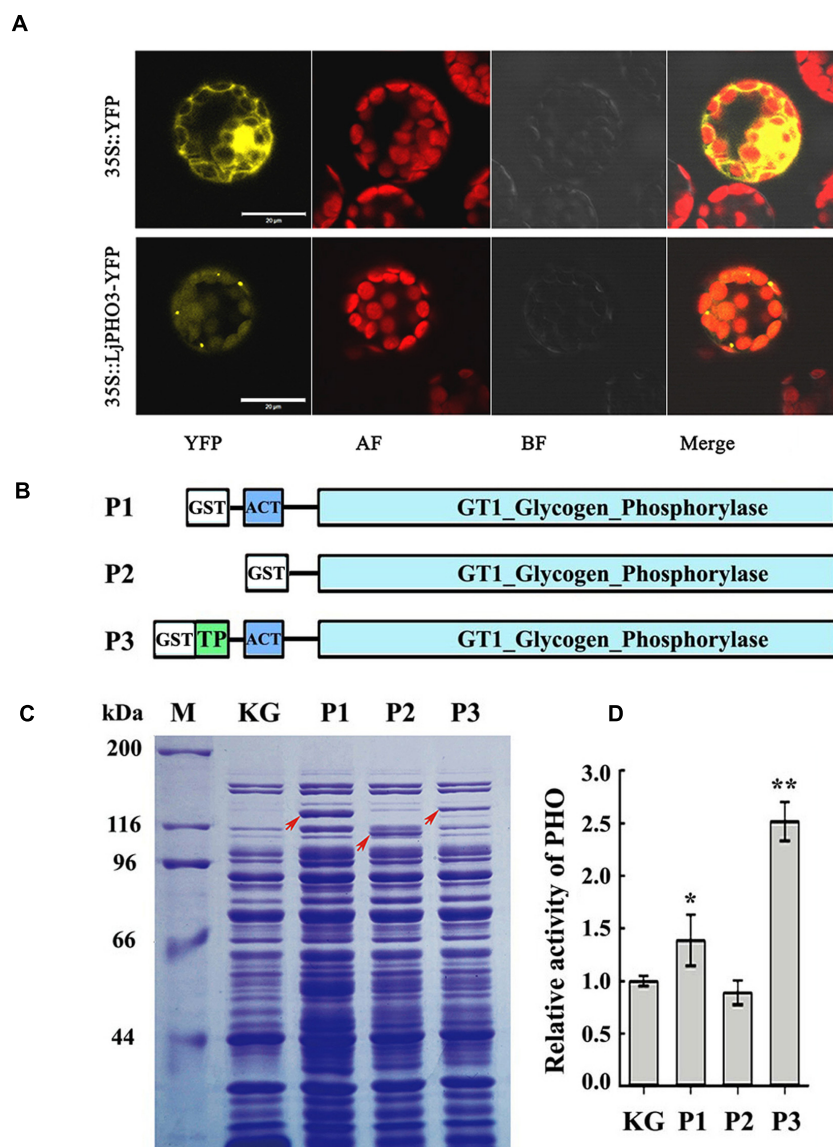


FIGURE 2 | Chloroplast localization of *LjPHO3*-eYFP and expression of *LjPHO3* in *Escherichia coli*. (A) Chloroplast localization of *LjPHO3*-eYFP. Protoplasts prepared from *Arabidopsis* leaves were used for gene transformation and observed by confocal laser scanning microscopy. Bar = 20 μ m; (B) Simplified graphical representation of expression of truncated *LjPHO3* genes in *E. coli*; (C) SDS-PAGE of bacterially expressed *LjPHO3* proteins. *E. coli* Rosetta was transformed with the native plasmid pGEX-KG or with reconstructed sequences (P1, P2, and P3), lysed, and the lysate was analyzed for PHO activity. A protein standard ladder (marker) is indicated on the right. The gel was stained with Coomassie Brilliant Blue R-250 dye; (D) Starch phosphorylase activity assay. Assays of the starch phosphorylase activity of PHO were carried out at 28°C, using oyster glycogen as substrate. The data represent averages of three replicates with mean standard deviations. (Duncan test: * $P < 0.05$; ** $P < 0.01$.)

laser scanning confocal microscopy. In P35S:LjPHO3-eYFP cells, eYFP fluorescence largely overlapped with the red fluorescence resulting from chlorophyll within chloroplasts (**Figure 2A**). This result suggested that LjPHO3 should be located in plastids. The strong fluorescent spots (**Figure 2A**) suggested the LjPHO3 proteins could also be deposited into protein bodies in the *Arabidopsis* cells.

LjPHO3 was cloned and expressed in *E. coli* Rosetta to determine whether this PHO gene encoded an authentic PHO enzyme. The recombinant forms of PHO3 so produced include P1 (Δ^{41} PHO, deletion of the putative TP domain sequences), P2 (Δ^{183} PHO, in which the putative TP domain and ACT domain sequences were deleted), and P3 (the full length coding domain; **Figure 2B**). After induction of expression by IPTG, starch phosphorylase activities of the total soluble protein in *E. coli* cells that contained recombinant P3 (full length) were on average 2.5-fold greater than the basal activity in *E. coli* cells containing the native

plasmid (**Figures 2C,D**). To further study the enzymatic characteristics of LjPHO3, we purified the recombinant PHO3 proteins of P1 and P3 using the GST-BIND™ Resin. Unfortunately, neither starch phosphorylase nor SS activity could be detected for any of the purified recombinant PHO3 proteins.

Overexpression of the *LjPHO3* Gene Decreased Starch Accumulation in *L. japonicus* Leaves

To investigate the function of PHO3 in *L. japonicus*, the gene was overexpressed in the MG-20 variety under the control of the CaMV 35S promoter. Three independent overexpression of *LjPHO3* transgenic lines (*LjPHO3*-OE1, 2, and 3) were established for use in these experiments (**Figure 3A**). Changes in the levels of *LjPHO3* transcripts were analyzed by qRT-PCR (**Figure 3B**).

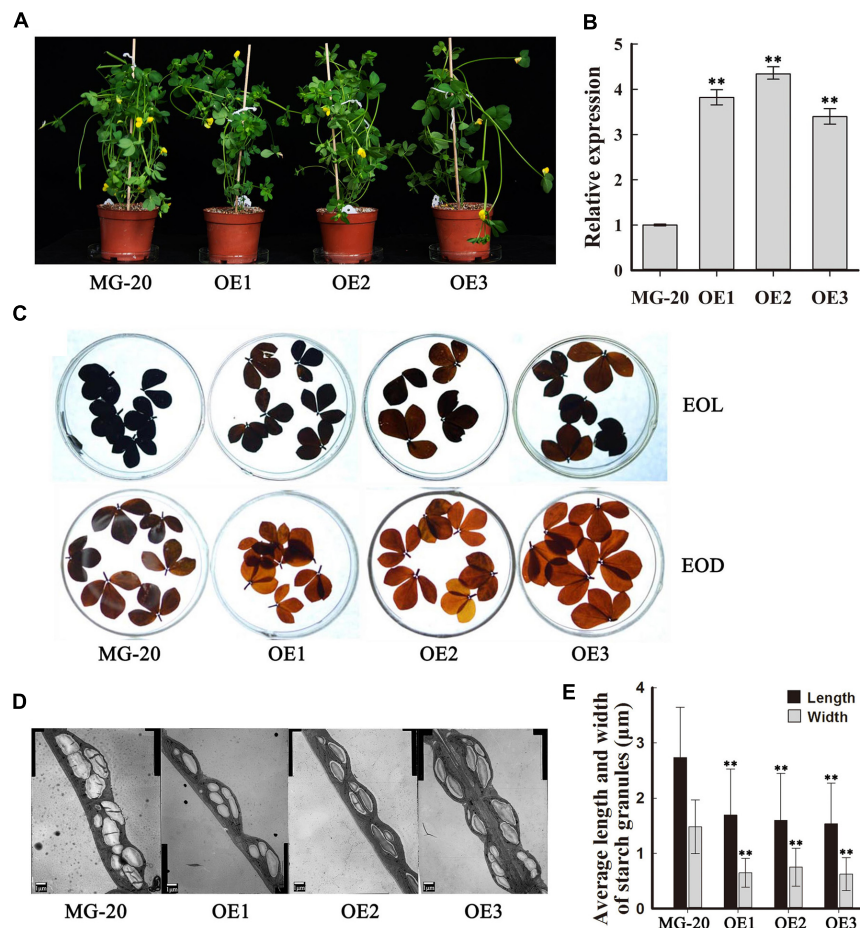


FIGURE 3 | Phenotype of transgenic plants carrying *LjPHO3* overexpression constructs. MG-20, wild-type *L. japonicus*; OE1, OE2, and OE3, three different transgenic lines. **(A)** Ten-week-old seedlings grown in a growth chamber; **(B)** Expression analysis of the *LjPHO3* gene in leaves of 4-week-old seedlings. Values represent means of $n = 6 \pm$ SD (Duncan test: * $P < 0.05$; ** $P < 0.01$); **(C)** Accumulation of starch in leaves of 10-week old seedlings. Leaves were heated in 80% (v/v) aqueous ethanol to remove pigments before staining with iodine (Lugol's solution). EOL, end of light; EOD, end of dark; **(D)** Electron microscopic observation of starch granules; **(E)** Average length and width of starch granules (SGs). Over 500 SGs were measured for each line (**Supplementary Figure S3**). Values represent means of $n \geq 500 \pm$ SD. (Duncan test: * $P < 0.05$; ** $P < 0.01$.)

Because of the predicted roles of PHO proteins in starch and oligosaccharide synthesis and/or degradation, we first measured the starch content in leaves of 10-week-old seedlings. The results showed that *LjPHO3-OE* leaves have an observable decrease in amounts of starch at both the end of the light period (EOL) and the end of the dark period (EOD; **Figure 3C**). Electron microscopic observation indicated that the starch granules at EOL were both shorter and narrower in *LjPHO3-OE* leaves than in wild-type leaves (**Figures 3D,E**; **Supplementary Figure S3B**).

A quantitative assay indicated that the starch content of *LjPHO3-OE* leaves was about 30% less than that of wild-type leaves (**Figure 4A**). On the other hand, the soluble sugar content was higher in *LjPHO3-OE* leaves than in wild-type leaves (**Figure 4B**). Next, we tested the activities of PHO and

three SS enzymes in *LjPHO3-OE* and wild-type leaves. The results showed that the *LjPHO3-OE* leaves have higher PHO starch phosphorolysis activity and SS activity, but lower AGPase activity, than the wild-type leaves (**Figures 4C–E**). No significant difference in BE activity (**Figure 4F**), nor changes in zymogram bands corresponding to PHO SS activity (**Supplementary Figure S3C**), were observed between the *LjPHO3-OE* and wild-type leaves.

Overexpression of the *LjPHO3* Gene Influences the Growth of *L. japonicus* Plants

Seeds of the *LjPHO3-OE* and wild-type lines were germinated, inoculated with *M. loti* MAFF303099, and grown on a vermiculite

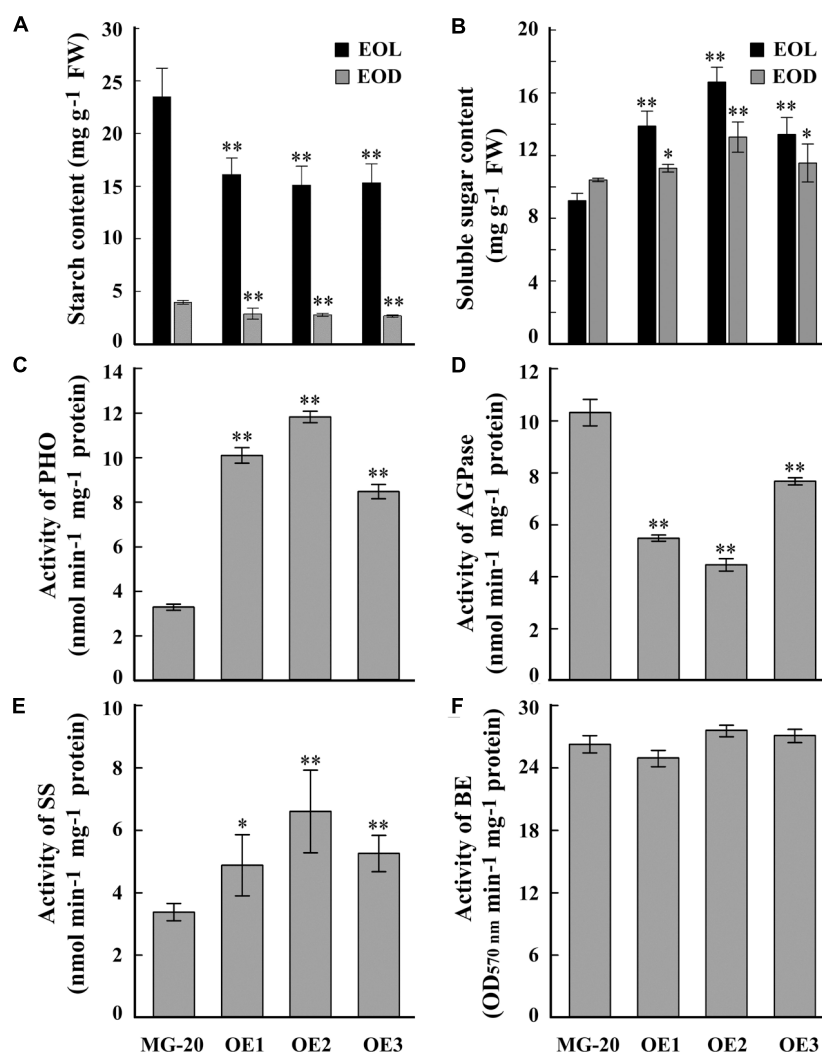


FIGURE 4 | Differences in biochemical indexes between wild-type leaves and *LjPHO3-OE* leaves. MG-20, wild-type *L. japonicus*; OE1, OE2, and OE3, three different transgenic lines. The leaves were taken from 10-week-old seedlings grown in a growth chamber. **(A,B)** Starch and soluble sugar contents. EOL, end of light; EOD, end of dark; **(C)** Starch phosphorolysis activity of PHO; **(D–F)** Activities of ADP-Glucose pyrophosphorylase (AGPase), **(D)** Starch synthase (SS), **(E)** Starch branching enzyme (BE), **(F)** The data represent averages of three biological replicates with mean standard deviations. (Duncan test: * $P < 0.05$; ** $P < 0.01$.)

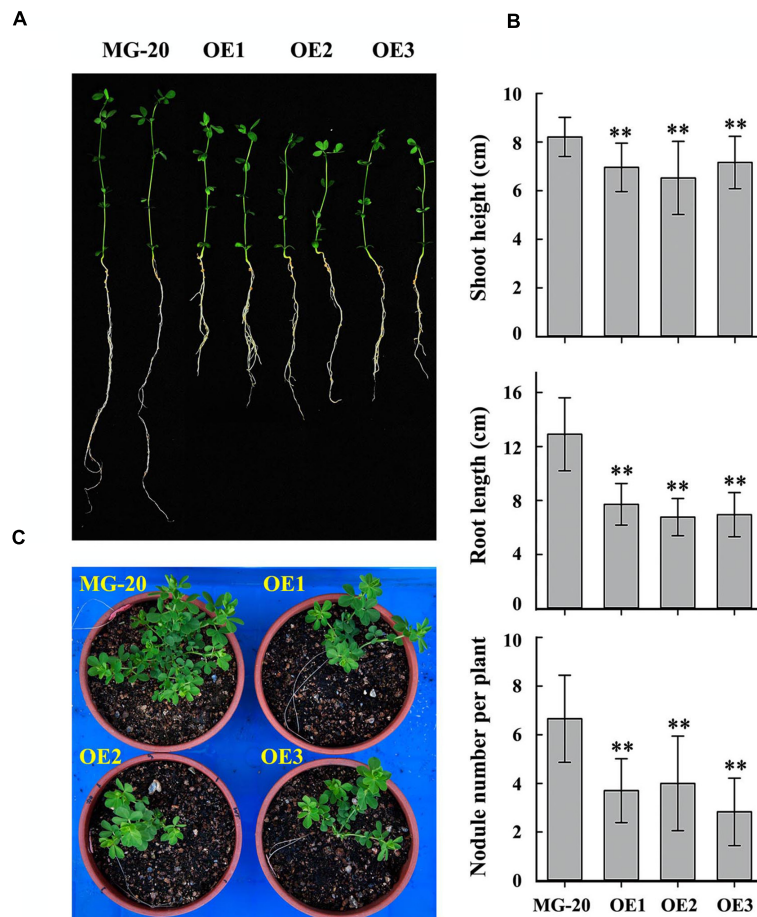


FIGURE 5 | Differences in plant size and nodule number between wild-type and *LjPHO3*-OE seedlings. MG-20, wild-type *L. japonicus*; OE1, OE2, and OE3, three different transgenic lines. **(A)** Four-week-old seedlings grown in a growth chamber; **(B)** Differences in shoot height, main root length, and nodule number between wild-type and *LjPHO3*-OE seedlings. The data represent averages of 30–40 seedlings with mean standard deviations (Duncan test: * $P < 0.05$; ** $P < 0.01$); **(C)** Ten-week-old seedlings grown in a greenhouse under natural daylight.

mixture for 4 weeks in a growth chamber (Figure 5A). The *LjPHO3*-OE seedlings were small with relatively shorter shoots and roots, and had fewer nodules, as compared to the wild-type seedlings (Figure 5B). At a later stage (10 weeks) under the greenhouse conditions, *LjPHO3*-OE seedlings were also small in comparison to the wild-type plants (Figure 5C), but there was no significant difference in flowering time. Their siliques were shorter, but their seeds were larger and heavier than those of wild-type plants (Supplementary Figure S4).

To study the influence of nitrogen on growth and starch accumulation in these plants, *LjPHO3*-OE and wild-type plants were germinated and grown on vermiculite (irrigated with Broughton and Dilworth nutrient solution containing 0, 5, and 10 mM KNO_3 , respectively) in transparent containers in a growth chamber. After 4 weeks, we observed that the *LjPHO3*-OE seedlings were small compared to the wild-type seedlings, especially under N-deficiency conditions (Figures 6A–C). The *LjPHO3*-OE leaves also had lower starch content (Figure 6D), but

higher soluble sugar content (Figure 6E), than wild-type leaves under all three N-supply conditions.

Overexpression of the *LjPHO3* Gene Decreased Pollen Fertility in *L. japonicus*

Compared to the wild-type, *LjPHO3*-OE plants displayed a 75% reduction in rate of seed set (Figures 7A,B). Wild-type pollen grains were round with ample cytoplasm, whereas many pollen grains from *LjPHO3*-OE plants were shrunk, irregular, and wizened, with scant cytoplasm (Figure 7C). The proportion of pollen grains from wild-type plants that became dark blue upon iodine staining was close to 100%, while the proportion from *LjPHO3*-OE plants was around 60% (Figure 7D). After incubation on medium for 30 min at 37°C the germination rate of pollen from *LjPHO3*-OE plants was around 60%, whereas the rate of germination for wild-type pollen was about 97% (Figures 7E,F). It is clear that overexpression of *LjPHO3* substantially reduces pollen fertility in *L. japonicus*.

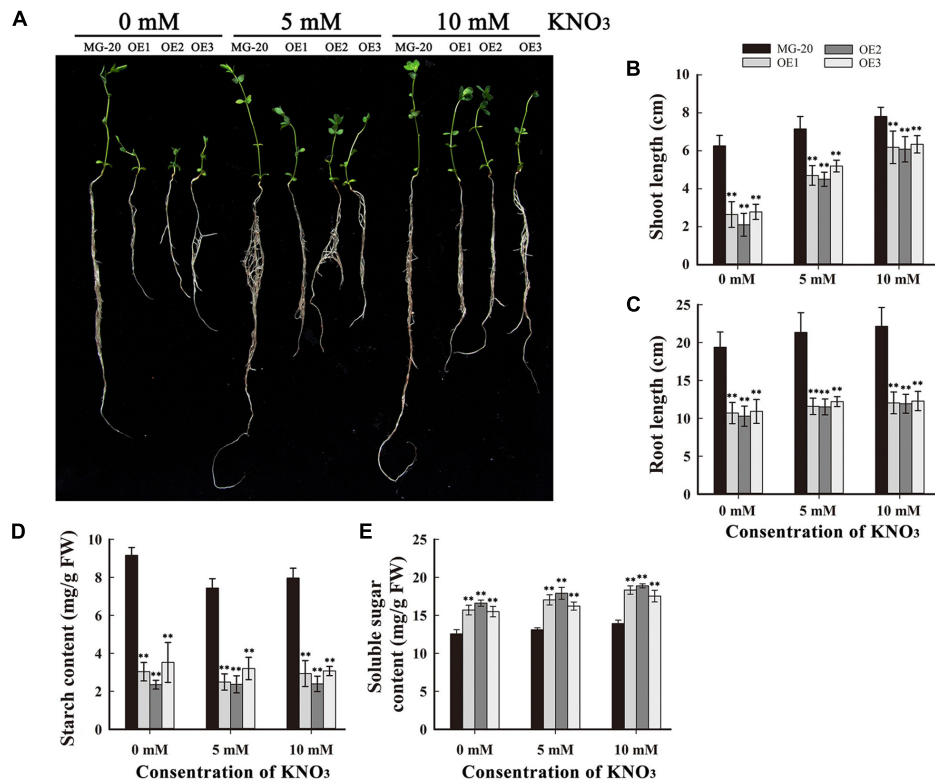


FIGURE 6 | Analysis of the effect of different levels of N-supply on the growth of *LjPHO3-OE* seedlings. MG-20, wild-type *L. japonicus*; OE1, OE2, and OE3, three different transgenic lines. (A) Four-week-old seedlings grown under different N-supply conditions (0, 5, and 10 mM KNO₃, respectively); (B,C) Shoot height (B) and main root length (C) of seedlings grown under different N-supply conditions; (D,E) Starch (D) and soluble sugar (E) contents of leaves at the end of the light period. The data represent averages of three biological replicates with mean standard deviations. (Duncan test: * $P < 0.05$; ** $P < 0.01$.)

DISCUSSION

Plastids are descended from a cyanobacterial symbiosis and starch metabolism genes probably coevolved following at least two rounds of whole genome duplication during the early period of plant evolution (Deschamps et al., 2008). It is therefore possible in principle for any plant lineage to have retained three isoforms of each starch metabolism protein. Only two starch phosphorylase isoforms, the plastidic starch phosphorylase (PHO1/PHOL) and the cytosolic starch phosphorylase (PHO2/PHOH), have previously been reported in higher plants (Deschamps et al., 2008). We detected a new starch phosphorylase family protein in some plants, which form a new subclade together with two *PHO* genes from the green algae *O. lucimarinus* and *C. reinhardtii* (*PHOB/STA4*) on the phylogenetic tree (Wu et al., 2015). In this study, we observed that the *PHO3* gene was present in the genome of *L. japonicus* (Supplementary Figure S1) and it participates in transitory starch metabolism in *L. japonicus* leaves. These results indicate the *PHO3* gene should be lost in some plant lineages during their evolution. The loss of genes encoding other SS enzymes was also observed in some plants. For examples, cruciferae plants such as *Arabidopsis* and *Brassica rapa* don't have the gene encoding BE I isoform in their genomes (Dumez et al., 2006). Many plants lost

the gene encoding SS VI isoform although its biological function is unknown (Wu et al., 2015). In these plants, the biological functions of these genes may be lost, or replaced by other genes during their evolution.

The PHOB/STA4 protein is located in the plastid in *C. reinhardtii* (Dauvillée et al., 2006). In *Arabidopsis* protoplasts, eYFP fluorescence of *LjPHO3-eYFP* largely overlapped with the red chlorophyll fluorescence within chloroplasts (Figure 2A), which indicates that *LjPHO3* is also a plastidic protein. Although the crude enzyme extract from *E. coli* cells containing the recombinant *LjPHO3* protein showed higher starch phosphorylase activity (Figure 2C), purified recombinant *LjPHO3* proteins did not show any SS or starch phosphorylase activities using the substrates glucose-1-phosphate, and/or maltose, maltotriose, maltoheptaose, glycogen, amylopectin, and amylose. The catalytic properties of the *LjPHO3* protein therefore remain to be determined.

In *C. reinhardtii*, the *sta4* mutants were found as mutants expressing a conditional low-starch high-amylose phenotype only in conditions of high polysaccharide synthesis (Libessart et al., 1995; Dauvillée et al., 2006). The authors speculated that the role of PHOB might be indirect through its involvement in a multienzyme complex (such as the BE-phosphorylase complex) selectively active during starch biosynthesis under high carbon

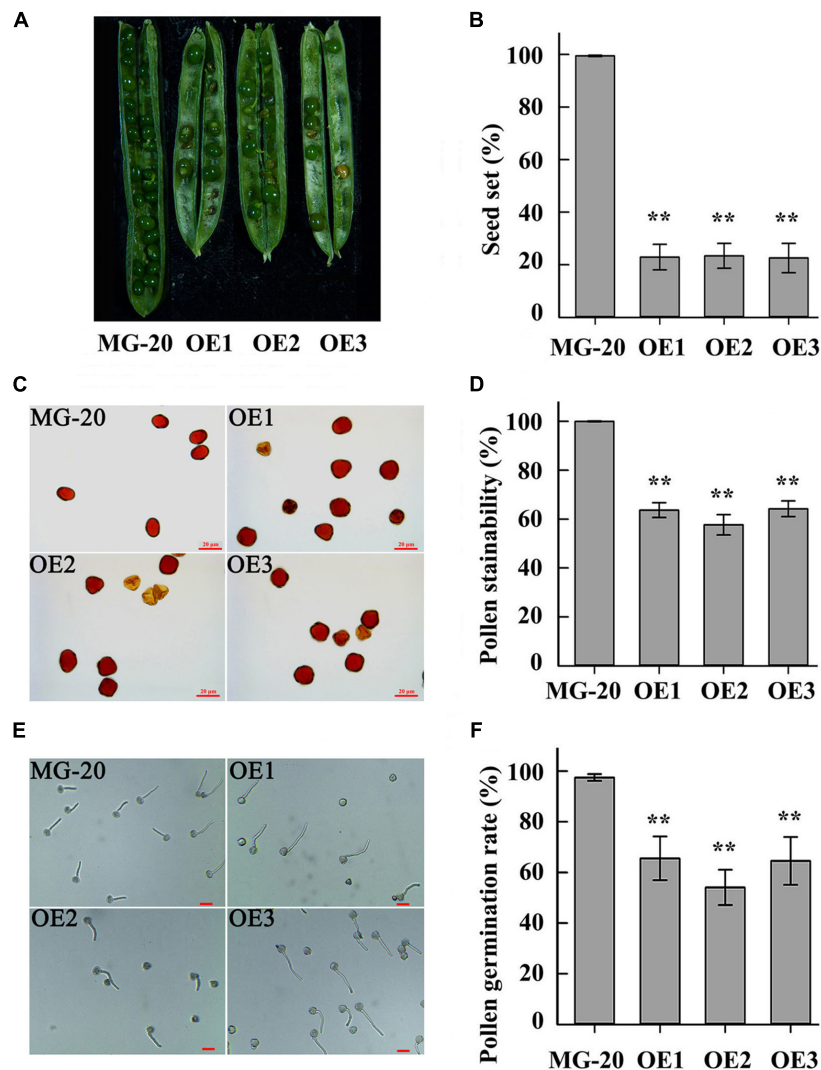


FIGURE 7 | Fertility changes in *LjPHO3*-OE plants. MG-20, wild-type *L. japonicus*; OE1, OE2, and OE3, three different transgenic lines. **(A)** Silique filling stage and seeds; **(B)** Rate of seed set. The data represent averages of over 50 siliques with mean standard deviations (Duncan test: * $P < 0.05$; ** $P < 0.01$.); **(C)** Morphology of pollen grains; **(D)** Stainability of pollen by iodine–potassium iodide solution; **(E)** *In vitro* pollen germination test. Scale bar = 20 μm; **(F)** Pollen germination rate. For each plant line, pollen grains from 5 to 6 anthers of each of 6 to 8 plants were analyzed. (Duncan test: * $P < 0.05$; ** $P < 0.01$.) Scale bar = 20 μm.

flux (Dauvillée et al., 2006). The majority of ACT domain-containing proteins appears to interact with amino acids and is involved in some aspect of regulation of amino acid metabolism (Grant, 2006). So, it could not exclude the possibility that one or more amino acids may be involved in the regulation of the activity of PHOB in starch turnover in *C. reinhardtii*.

Like the *C. reinhardtii* PHOB/STA4, the *LjPHO3* protein has an ACT domain [cl09141] in the N-terminus, but it lacks the L78 domain which is present in PHO1 proteins (Figure 1A; Supplementary Figure S2). In this study, we observed that the *LjPHO3* proteins appeared to play a negative role in starch accumulation in *L. japonicus* leaves. Overexpression of *LjPHO3* in *L. japonicus* resulted in a significant decrease in amounts of starch and sizes of starch granules but an increased amount of

soluble sugars in leaves of the transgenic plants (Figures 3C and 4A,B; Supplementary Figure S3). This effect was not affected by nitrogen supplying levels (Figures 6D,E), but the growth of *LjPHO3*-OE seedlings was seriously retarded under N-deficiency conditions (Figure 6A). These results implied that the increased activity of phosphorylase (Figure 4C) increased the phosphorolysis of the transitory starches in the *LjPHO3*-OE leaves. The increase of SS activity (Figure 4E) may be induced by the elevated content of sugars in *LjPHO3*-OE leaves (Figure 4B). In rice and *Arabidopsis*, the expression levels of SS genes, but not *AGPase* genes those function in leaves have been reported to be upregulated by higher sugars in leaves (Dian et al., 2003, 2005; Niittylä et al., 2004; Akihiro et al., 2005). Reduction in starch accumulation may elevate sugar export from chloroplasts further

led to the increase of Pi content in chloroplasts. In leaves of higher plants, the activity of AGPase is mainly modified at protein level, such as regulated by 3-phosphoglycerate (activator) and inorganic orthophosphate (inhibitor), and also redox modification (Ballicora et al., 2004). The reduction of AGPase activity in *LjPHO3-OE* leaves (**Figure 4D**) may be due to changes in Pi content and sugar (-P) pool in chloroplasts. The detail changes in photosynthetic parameters, sugar/sugar-P composition and other compounds between the wild-type and transgenic leaves remained to be determined next. It has been reported that transitory starch is produced in the chloroplast of leaves during the day, and degrades to support metabolism and growth at night in many plants (Smith and Stitt, 2007; Vriet et al., 2010). In consequence, the functional transitory starch turnover is crucial for normal growth of plants. The smaller size of *LjPHO3-OE* seedlings may be due to an insufficiency of sugars supplied by the leaves at night (**Figures 5A,C and 6A**). The reduction in pollen fertility (**Figure 7**) may be due to the decreased in deposition of starch in developing pollen grains of *LjPHO3-OE* plants. Alternatively, it may be because of the insufficient amount of sugars supplied by the leaves. As a result of so many of the pollen grains being sterile, seed set was reduced in *LjPHO3-OE* plants (**Figure 7**).

Considering the phenotype of the overexpression lines it was logical to test the effect of *LjPHO3* knock-down. Two different approaches were used, including RNA interference lines (MG-20) and two LTR retrotransposon insertion lines (Gifu; DK02: No.30006161 and DK07: No.30056666 from the LORE1 insertion mutant resource¹⁴). Sadly, *LjPHO3* expression was not reduced significantly in the interference lines and those plants did not exhibit related phenotypes. Furthermore, the two tested insertion lines, were not knock-out mutants, as full length transcript cDNA was detected in homozygote plants (data not shown). It would be interesting to test the newly available insertion lines to conclude on the effect of *LjPHO3* knock-out.

Starch synthesis or starch degradation? The plastidial PHO1 may play different physiological role between species, growing conditions and tissues, and one or more unknown factors were involved in regulating the action (Zeeman et al., 2004; Satoh et al., 2008; Lin et al., 2012; Streb and Zeeman, 2012; Malinova et al., 2014). The same assumption could also apply to the physiological role of the PHO3 protein. In *C. reinhardtii*, PHOB/PHO3 may play a role in storage starch biosynthesis in conditions of nitrogen starvation (Dauvillée et al., 2006). In *L. japonicus*, PHO3 should participate in starch degradation in leaves based on the present results. On the other hand, overexpression of the *LjPHO3* gene in rice did not significantly affect the starch content of either leaves or endosperms, or the rate of seed set rate in transgenic rice plants (data not shown). These results imply that one or more additional factor (s) in *L. japonicus* plants are required for the PHO3 protein to be active in regulating starch accumulation. They could also explain why the purified recombinant PHO3 proteins showed no SS or starch phosphorylase activity *in vitro*.

¹⁴ <http://lotus.au.dk>

CONCLUSION

Lotus japonicus has a gene encoding the PHO3 isoform. The *LjPHO3* protein lacks the L78 domain but has an ACT domain, and it is located in the chloroplast. Overexpression of *LjPHO3* in *L. japonicus* results in a reduction in starch content in leaves. The reduction in starch deposition has a major impact on plant growth, pollen fertility, and rate of seed set rate in the transgenic plants. However, the catalytic properties of the *LjPHO3* protein remain to be further studied. To our knowledge, the results presented here represent the first data on the function of the PHO3 protein subfamily in higher plants.

AUTHOR CONTRIBUTIONS

The research was designed by HJ, GW, YC, ML, and PW. The experiments were performed by SQ, YT, and the data were analyzed by SQ. The manuscript was written by SQ.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (31070227) and the External Cooperation Program of BIC, Chinese Academy of Sciences (151644KYSB20130054).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01315>

FIGURE S1 | Neighbor-joining unrooted tree. Bootstrap values were calculated for 100 replicates, and values are indicated at the corresponding nodes. The branch length corresponding to the number of substitutions per site is given and the database accession numbers of sequences are indicated in brackets.

FIGURE S2 | Comparison of the deduced amino acid sequences of PHO proteins from *Lotus japonicus* and *Chlamydomonas*. Conserved amino acids are indicated by shaded squares.

FIGURE S3 | Phenotype of starch granules and starch phosphorylase isoforms in wild-type leaves and *LjPHO3-OE* leaves. (A) Electron microscopic observation of starch granules. More chloroplasts lacked starch granules in *LjPHO3-OE* plants than in wild-type plants; (B) Distribution of the lengths and widths of measured starch granules; (C) Zymogram analysis of starch synthesis activities of the PHO enzymes. Soluble proteins from crude extracts of leaves were subjected to native PAGE [7.5% (w/v) acrylamide slab gel containing 0.5% (w/v) oyster glycogen] and stained to reveal phosphorylase isoforms.

FIGURE S4 | Differences in silique length and seed size between MG-20 and *LjPHO3-OE* seedlings. (A) and (B) Difference in siliques length between MG-20 and *LjPHO3-OE* seedlings; (C) Seed size; (D) The 1000-seed weight. Values represent means of $n = 6 \pm \text{SD}$. (Duncan test: * $P < 0.05$; ** $P < 0.01$.)

REFERENCES

- Akihiro, T., Mizuno, K., and Fujimura, T. (2005). Gene expression of ADP-glucose pyrophosphorylase and starch contents in rice cultured cells are cooperatively regulated by sucrose and ABA. *Plant Cell Physiol.* 46, 937–946. doi: 10.1093/pcp/pci101
- Ballicora, M. A., Iglesias, A. A., and Preiss, J. (2004). ADP-Glucose pyrophosphorylase: a regulatory enzyme for plant starch synthesis. *Photosynth. Res.* 79, 1–24. doi: 10.1023/B:PRES.0000011916.67519.58
- Chen, H. M., Chang, S. C., Wu, C. C., Cuo, T. S., Wu, J. S., and Juang, R. H. (2002). Regulation of the catalytic behaviour of L-form starch phosphorylase from sweet potato roots by proteolysis. *Physiol. Plant* 114, 506–515. doi: 10.1034/j.1399-3054.2002.1140402.x
- Chen, Y. P., Chen, W., Li, X. L., Jiang, H. W., Wu, P. Z., Xia, K. F., et al. (2014). Knockdown of LjIPT3 influences nodule development in *Lotus japonicus*. *Plant Cell Physiol.* 55, 183–193. doi: 10.1093/pcp/pct171
- Dauvillée, D., Chochois, V., Steup, M., Haebel, S., Eckermann, N., Ritte, G., et al. (2006). Plastidial phosphorylase is required for normal starch synthesis in *Chlamydomonas reinhardtii*. *Plant J.* 48, 274–285. doi: 10.1111/j.1365-313X.2006.02870.x
- Deschamps, P., Moreau, H., Worden, A. Z., Dauvillée, D., and Ball, S. G. (2008). Early gene duplication within Chloroplastida and its correspondence with relocation of starch metabolism to chloroplasts. *Genetics* 178, 2373–2387. doi: 10.1534/genetics.108.087205
- Dian, W., Jiang, H., Chen, Q., Liu, F., and Wu, P. (2003). Cloning and characterization of the granule-bound starch synthase II gene in rice, gene expression is regulated by the nitrogen level, sugar and circadian rhythm. *Planta* 218, 261–268. doi: 10.1007/s00425-003-1101-9
- Dian, W., Jiang, H., and Wu, P. (2005). Evolution and expression analysis of starch synthase III and IV in rice. *J. Exp. Bot.* 56, 623–632. doi: 10.1093/jxb/eri065
- DuBois, M., Gilles, K., Hamilton, J. K., Rebers, P. A., and Smith, F. (1951). A colorimetric method for the determination of sugars. *Nature* 168, 167–167. doi: 10.1038/168167a0
- Dumez, S., Wattedled, F., Dauvillée, D., Delvalle, D., Planhot, V., Ball, S. G., et al. (2006). Mutants of *Arabidopsis* lacking starch branching enzyme II substitute plastidial starch synthesis by cytoplasmic maltose accumulation. *Plant Cell* 18, 2694–2709. doi: 10.1105/tpc.105.037671
- Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/ps.8.5.978
- Fettke, J., Chia, T., Eckermann, N., Smith, A., and Steup, M. (2006). A transglucosidase necessary for starch degradation and maltose metabolism in leaves at night acts on cytosolic heteroglycans (SHG). *Plant J.* 46, 668–684. doi: 10.1111/j.1365-313X.2006.02732.x
- Fettke, J., Poeste, S., Eckermann, N., Tiessen, A., Pauly, M., Geigenberger, P., et al. (2005). Analysis of cytosolic heteroglycans from leaves of transgenic potato (*Solanum tuberosum* L.) plants that under- or overexpress the Pho 2 phosphorylase isozyme. *Plant Cell Physiol.* 46, 1987–2004. doi: 10.1093/pcp/pci214
- Grant, G. A. (2006). The ACT domain: a small molecule binding domain and its role as a common regulatory element. *J. Biol. Chem.* 281, 33825–33829. doi: 10.1074/jbc.R600024200
- Hwang, S. K., Singh, S., Kahir, B., Satoh, H., and Okita, T. W. (2016). The plastidial starch phosphorylase from rice endosperm: catalytic properties at low temperature. *Planta* 243, 999–1009. doi: 10.1007/s00425-015-2461-7
- Jiang, H., Dian, W., and Wu, P. (2003). Effect of high temperature on fine structure of amylopectin in rice endosperm by reducing the activity of the starch branching enzyme. *Phytochemistry* 63, 53–59. doi: 10.1016/S0031-9422(03)00005-0
- Jiang, H. W., Li, M. R., Liang, N. T., Yan, H. B., Wei, Y. B., Xu, X. L., et al. (2007). Molecular cloning and function analysis of the stay green gene in rice. *Plant J.* 52, 197–209. doi: 10.1111/j.1365-313X.2007.03221.x
- Kruger, N. J., and ap Rees, T. (1983). Properties of α -glucan phosphorylase from pea chloroplasts. *Phytochemistry* 22, 1891–1898. doi: 10.1016/0031-9422(83)80007-7
- Libessart, N., Maddelein, M. L., Koornhuyse, N., Decq, A., Delrue, B., Mouille, G., et al. (1995). Storage, photosynthesis, and growth: the conditional nature of mutations affecting starch synthesis and structure in *Chlamydomonas*. *Plant Cell* 7, 1117–1127. doi: 10.1105/tpc.7.8.1117
- Lin, Y. C., Chen, H. M., Chou, I. M., Chen, A. N., Chen, C. P., Young, G. H., et al. (2012). Plastidial starch phosphorylase in sweet potato roots is proteolytically modified by protein-protein interaction with the 20S proteasome. *PLoS ONE* 7:e35336. doi: 10.1371/journal.pone.0035336
- Lu, Y., Steichen, J. M., Yao, J., and Sharkey, T. D. (2006). The role of cytosolic α -glucan phosphorylase in maltose metabolism and the comparison of amyloamylase in *Arabidopsis* and *Escherichia coli*. *Plant Physiol.* 142, 878–889. doi: 10.1104/pp.106.086850
- Malinova, I., Mahlow, S., Alseekh, S., Orawetz, T., Fernie, A. R., Baumann, O., et al. (2014). Double knockout mutants of *Arabidopsis* grown under normal conditions reveal that the plastidial phosphorylase isozyme participates in transitory starch metabolism. *Plant Physiol.* 164, 907–921. doi: 10.1104/pp.113.227843
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., et al. (2015). CDD, NCBI's conserved domain database. *Nucleic Acids Res.* 43, D222–D226. doi: 10.1093/nar/gku1221
- Mori, H., Tanizawa, K., and Fukui, T. (1993). A chimeric α -glucan phosphorylase of plant type L and H isozymes. Functional role of 78-residue insertion in type L isozyme. *J. Biol. Chem.* 268, 5574–5581.
- Niittylä, T., Messerli, G., Trevisan, M., Chen, J., Smith, A. M., and Zeeman, S. C. (2004). A previously unknown maltose transporter essential for starch degradation in leaves. *Science* 303, 87–89. doi: 10.1126/science.1091811
- Nishi, A., Nakamura, Y., Tanaka, N., and Satoh, H. (2001). Biochemical and genetic analysis of the effects of amylose-extender mutation in rice endosperm. *Plant Physiol.* 127, 459–472. doi: 10.1104/pp.010127
- Satoh, H., Shibahara, K., Tokunaga, T., Nishi, A., Tasaki, M., Hwang, S. K., et al. (2008). Mutation of the plastidial α -glucan phosphorylase gene in rice affects the synthesis and structure of starch in the endosperm. *Plant Cell* 20, 1833–1849. doi: 10.1105/tpc.107.054007
- Smith, A. M., and Stitt, M. (2007). Coordination of carbon supply and plant growth. *Plant Cell Environ.* 30, 1126–1149. doi: 10.1111/j.1365-3040.2007.01708.x
- Streb, S., and Zeeman, S. C. (2012). Starch metabolism in *Arabidopsis*. *Arabidopsis Book* 10:e0160. doi: 10.1199/tab.0160
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4, molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Vriet, C., Welham, T., Brachmann, A., Pike, M., Pike, J., Perry, J., et al. (2010). A Suite of *Lotus japonicus* starch mutants reveals both conserved and novel features of starch metabolism. *Plant Physiol.* 154, 643–655. doi: 10.1104/pp.110.161844
- Wirtz, W., Stitt, M., and Heldt, H. W. (1980). Enzymic determination of metabolites in the subcellular compartments of spinach protoplasts. *Plant Physiol.* 66, 187–193. doi: 10.1104/pp.66.1.187
- Wu, P. Z., Zhou, C. P., Cheng, S. F., Wu, Z. Y., Lu, W. J., Han, J., et al. (2015). Integrated genome sequence and linkage map of physic nut (*Jatropha curcas* L.), a biodiesel plant. *Plant J.* 81, 810–821. doi: 10.1111/tjp.12761
- Yoo, S. D., Cho, Y. H., and Sheen, J. (2007). *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. *Nat. Protoc.* 2, 1565–1572. doi: 10.1038/nprot.2007.199
- Zeeman, S. C., Thorneycroft, D., Schupp, N., Chapple, A., Weck, M., Dunstan, H., et al. (2004). Plastidial α -glucan phosphorylase is not required for starch degradation in *Arabidopsis* leaves but has a role in the tolerance of abiotic stress. *Plant Physiol.* 135, 849–858. doi: 10.1104/pp.103.032631

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer AT and handling Editor declared their shared affiliation, and the handling Editor states that the process nevertheless met the standards of a fair and objective review.

Copyright © 2016 Qin, Tang, Chen, Wu, Li, Wu and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Arbuscular Mycorrhiza Stimulates Biological Nitrogen Fixation in Two *Medicago* spp. through Improved Phosphorus Acquisition

David Püschel^{1,2*}, Martina Janoušková¹, Alena Voříšková¹, Hana Gryndlerová², Miroslav Vosátka¹ and Jan Jansa²

¹ Department of Mycorrhizal Symbioses, Institute of Botany, Czech Academy of Sciences, Průhonice, Czechia, ² Laboratory of Fungal Biology, Institute of Microbiology, Czech Academy of Sciences, Prague, Czechia

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Consejo Superior de Investigaciones
Científicas (CSIC), Spain

Reviewed by:

Iver Jakobsen,
University of Copenhagen, Denmark
Stephan Unger,
Bielefeld University, Germany

*Correspondence:

David Püschel
david.puschel@ibot.cas.cz

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 29 November 2016

Accepted: 07 March 2017

Published: 27 March 2017

Citation:

Püschel D, Janoušková M,
Voříšková A, Gryndlerová H,
Vosátka M and Jansa J (2017)
Arbuscular Mycorrhiza Stimulates
Biological Nitrogen Fixation in Two
Medicago spp. through Improved
Phosphorus Acquisition.
Front. Plant Sci. 8:390.
doi: 10.3389/fpls.2017.00390

Legumes establish root symbioses with rhizobia that provide plants with nitrogen (N) through biological N fixation (BNF), as well as with arbuscular mycorrhizal (AM) fungi that mediate improved plant phosphorus (P) uptake. Such complex relationships complicate our understanding of nutrient acquisition by legumes and how they reward their symbiotic partners with carbon along gradients of environmental conditions. In order to disentangle the interplay between BNF and AM symbioses in two *Medicago* species (*Medicago truncatula* and *M. sativa*) along a P-fertilization gradient, we conducted a pot experiment where the rhizobia-treated plants were either inoculated or not inoculated with AM fungus *Rhizophagus irregularis* 'PH5' and grown in two nutrient-poor substrates subjected to one of three different P-supply levels. Throughout the experiment, all plants were fertilized with ¹⁵N-enriched liquid N-fertilizer to allow for assessment of BNF efficiency in terms of the fraction of N in the plants derived from the BNF (%N_{BNF}). We hypothesized (1) higher %N_{BNF} coinciding with higher P supply, and (2) higher %N_{BNF} in mycorrhizal as compared to non-mycorrhizal plants under P deficiency due to mycorrhiza-mediated improvement in P nutrition. We found a strongly positive correlation between total plant P content and %N_{BNF}, clearly documenting the importance of plant P nutrition for BNF efficiency. The AM symbiosis generally improved P uptake by plants and considerably stimulated the efficiency of BNF under low P availability (below 10 mg kg⁻¹ water extractable P). Under high P availability (above 10 mg kg⁻¹ water extractable P), the AM symbiosis brought no further benefits to the plants with respect to P nutrition even as the effects of P availability on N acquisition via BNF were further modulated by the environmental context (plant and substrate combinations). As a response to elevated P availability in the substrate, the extent of root length colonization by AM fungi was reduced, the turning points occurring at about 8 and 10 mg kg⁻¹ water extractable P for *M. sativa* and *M. truncatula*, respectively. Our results indicated competition for limited C resource between the two kinds of microsymbionts and thus degradation of AM symbiotic functioning under ample P supply.

Keywords: legumes, root symbioses, rhizobia, arbuscular mycorrhiza, nitrogen acquisition, phosphorus uptake, competition, synergies

INTRODUCTION

Legumes form two different types of root symbioses with soil microorganisms. Rhizobial symbiosis, exclusive to legumes, is established with soil diazotrophic bacteria that induce formation of nodules in host plants' roots. Rhizobia fix atmospheric dinitrogen (N_2) and provide it to the plants in the ammonium form that can easily be assimilated by the plant. Biological nitrogen fixation (BNF) thus contributes significantly to the nitrogen (N) budget of legumes. Its share in total N uptake by the plants is estimated to reach as high as 65–95% (Bolger et al., 1995). Arbuscular mycorrhizal (AM) symbiosis is by far more widespread among plant taxa. This association is established between the majority of terrestrial vascular plants and AM fungi from the phylum Glomeromycota (Smith and Read, 2008). AM fungi colonize plant roots and then their hyphae radiate into the surrounding soil, creating extensive networks of mycelium reaching to soil volume up to two orders of magnitude greater than what is accessible by plants alone (Raven and Edwards, 2001) and thus well beyond the depletion zone of the roots of poorly mobile nutrients such as phosphorus (P). The pivotal role of AM symbiosis occurs in enhancing plants' uptake of such poorly mobile nutrients as P and/or zinc (Jansa et al., 2003a, 2011; Kiers et al., 2011).

In both rhizobial and AM symbioses, plants reward their microbial partners with photosynthetically assimilated carbon (C). The flows of C, N, and P in plants hosting both symbionts thus becomes rather complex (Figure 1). Each symbiosis may consume around 3–20% of recently fixed C to maintain the growth and activity and to build up energy reserves of the participating microbes (Jakobsen and Rosendahl, 1990; Kaschuk et al., 2009; Slavíková et al., 2016). The plants can partly compensate for C needs of their symbionts by increased CO_2 assimilation (Paul and Kucey, 1981), either due to C sink stimulation or indirectly through the nutritional benefits received from the symbioses (Kaschuk et al., 2009; Řezáčová et al., 2017).

Although the functioning of either of these symbioses alone has been studied in depth through past decades, their interaction remains insufficiently explored. Synergistic effects on plants of rhizobial and mycorrhizal symbioses have been described (e.g., Kaschuk et al., 2010; Larimer et al., 2014; van der Heijden et al., 2016), but the interaction of the two symbionts may also reduce plant growth (e.g., Bethlenfalvay et al., 1982; Ballhorn et al., 2016). As pointed out in a review by Larimer et al. (2010), there is a need for more experimental studies relating the interaction of the symbionts to abiotic conditions because nutrient availability and other environmental factors may influence the outcome. For example, Saia et al. (2014) observed that AM symbiosis enhanced BNF and total plant biomass under drought stress but not under water-sufficient conditions. Ballhorn et al. (2016) reported interactive effects of AM and rhizobial symbioses depending on light availability. Surprisingly, however, there is only limited and inconclusive information on how the interaction of the two symbionts changes along the P-availability gradients (see Bethlenfalvay et al., 1982; Larimer et al., 2014).

Maintenance of rhizobial symbiosis imposes great P demand on the host plants (Jakobsen, 1985). This is because nodules

have high sink-strength for P, probably due to considerable nitrogenase demand for ATP and because P concentration in microbial tissue is substantially higher than in plant cells (Jakobsen, 1985). Under low P availability to the plants, the efficiency of BNF thus often decreases (Kleinert et al., 2014). This effect is thought to be merely indirect, through intensifying P deficiency of the host plant and, as a consequence, impairing the photosynthetic capacity of the host plant, and not directly affecting nodule formation or function (Jakobsen, 1985). As AM fungi usually improve their host plants' P status under low P availability, AM symbiosis is expected to support rhizobial activity and increase BNF. At high P availability, AM symbiosis usually does not further improve the host plant's P budget (Smith and Read, 2008) and is therefore unlikely to increase BNF through improved P nutrition. In contrast, the host plant may become C-limited under P-sufficient conditions (Johnson et al., 1997) with the consequence that synergy of the two symbionts changes to antagonism as the two compete for C as the system-limiting resource (Bethlenfalvay et al., 1982; Reinhard et al., 1993).

To achieve a better understanding of the role of P availability in the functional interplay between rhizobial and mycorrhizal symbioses, a factorial pot experiment with two soil types and three P-fertilization levels was conducted using mycorrhizal and non-mycorrhizal individuals of the species *Medicago truncatula* and *M. sativa* inoculated with their compatible rhizobia. The plants were grown in substrates differing in pH and P availability, and fertilized with ^{15}N -labeled ammonium nitrate to allow assessment of BNF contribution to the plants' N uptake. We hypothesized that (1) the efficiency of BNF would positively correlate with P nutrition of the plants, and (2) under low P availability in the substrate, mycorrhizal plants would acquire relatively more N from BNF than would non-mycorrhizal plants due to the functional synergy between the two symbioses.

MATERIALS AND METHODS

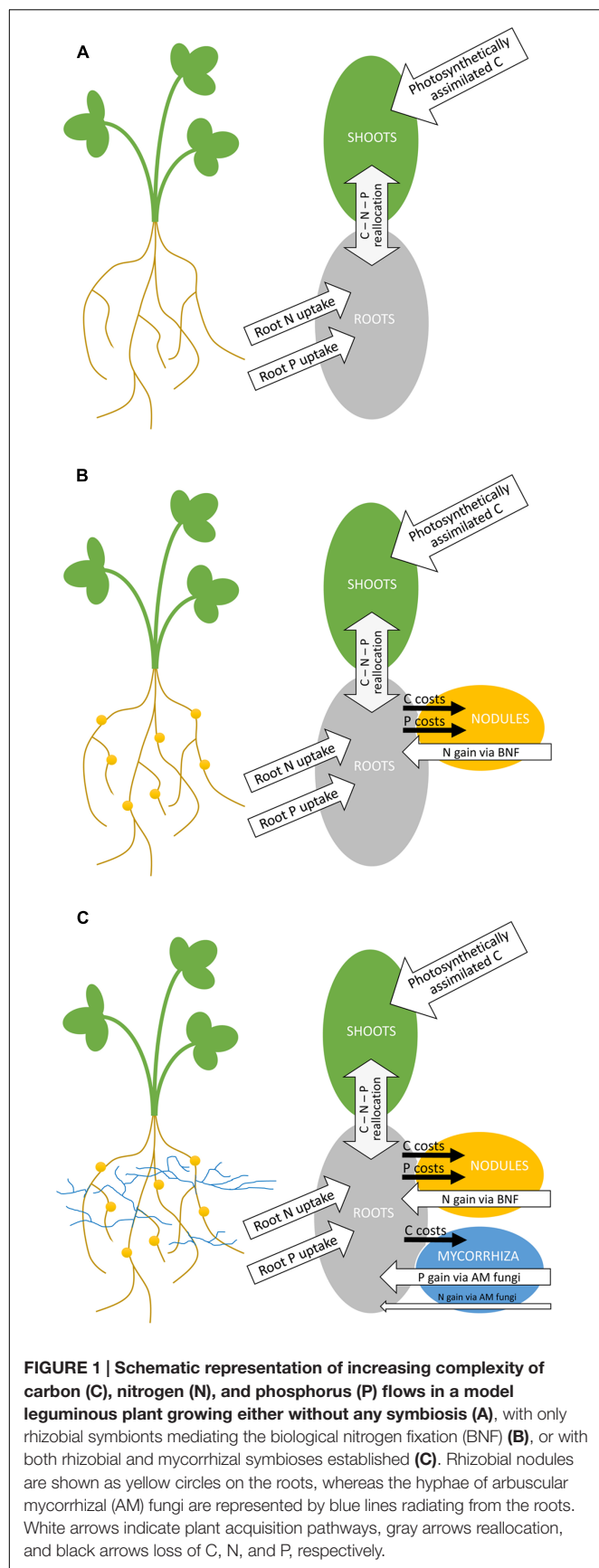
Experimental Design

In a greenhouse pot experiment, two model plant species, *M. truncatula* and *M. sativa*, were planted in two different substrates amended or not with mineral P-fertilizer to reach three levels of P availability for each of the substrates. All plants were inoculated with rhizobia compatible with the respective host plant species. Half of the plants were further inoculated with AM fungal isolate *Rhizophagus irregularis* 'PH5', whereas the other plants grew without the AM fungus. The experiment was conducted in a fully factorial experimental design with five biological replicates per treatment, and thus it comprised 120 pots.

Substrate and Initial P-Fertilization

A mixture (1:1, v:v) of autoclaved (at 121°C for 30 min) quartz sand (grain size < 4 mm) and autoclaved zeolite (grain size 1–2.5 mm; Zeopol s.r.o., Břeclav, Czech Republic)¹ provided the

¹<http://www.zeopol.com>



basis of the substrate used in this study. To this sand–zeolite mixture, 10% (of final volume) of γ -irradiated (>25 kGy) soil was added. Two soils of different origins and with different physicochemical properties were used. The first soil originated from Litoměřice, Czech Republic (GPS coordinates 50.532°N, 14.110°E) and the second soil was obtained from Tänikon, Switzerland (47.489°N, 8.919°E). The two soils differed in pH and calcium content and thus were assumed to have different P saturation kinetics, effectively resulting in a 6-point P-availability gradient obtained as a combination of 2 substrates and 3 P-supply levels (see **Table 1** for selected physicochemical properties of the different substrates).

The substrates (further referred to as “LT” or “Tän” substrate depending on the identity of the soil component) were filled into tall, 2 L plastic pots (11 cm \times 11 cm \times 20 cm). First, the bottom third of each pot’s volume was filled. Prior to filling the upper two thirds of the pots, the respective volume of the substrate was subjected to initial fertilization and/or mycorrhizal inoculation, as required by the specific experimental treatment (details described below).

To prevent plant growth limitation due to lack of potassium (K), magnesium (Mg), and/or calcium (Ca), these nutrients were uniformly added into all pots as initial fertilization of the substrate. The doses of 60 mg of K, 30 mg of Mg, and 30 mg of Ca (per pot) were provided by means of two separate nutrient solutions that were prepared by dissolving either 13.372 g of K_2SO_4 together with 30.423 g of $MgSO_4 \cdot 7H_2O$, or 11.005 g of $CaCl_2 \cdot 2H_2O$ per 1 L of distilled water. Both solutions were applied in doses of 10 mL per pot and thoroughly mixed into the upper two thirds of the substrate filled into each pot.

The gradient of P supply comprised three levels, hereafter referred to as “P0,” “P10,” and “P40,” with either 0, 10, or 40 mg of P added to the pots (**Table 1**). This was achieved by applying

TABLE 1 | Selected physico-chemical properties of sand–zeolite (1:1, v:v) substrates with 10% volumetric content of either Tänikon (Tän) or Litoměřice (LT) soil, and supplemented with either 0, 10, or 40 mg of P per pot (P0, P10, and P40, respectively).

Substrate	pH H_2O	C [%]	N [%]	Available P [mg kg^{-1}]	Total P [mg kg^{-1}]
Tän P0	7.52	0.186	0.012	2.06	73.06
Tän P10	7.51	0.230	0.018	5.00	85.09
Tän P40	7.43	0.226	0.015	12.35	125.96
LT P0	8.28	0.260	0.010	3.03	108.52
LT P10	8.19	0.307	0.012	5.49	158.68
LT P40	8.33	0.214	0.008	8.32	161.81

The pH was measured in 1:2.5 (w:v) water slurry following 30 min shaking at room temperature by pH meter FE20 (Mettler Toledo). Available (water extractable) phosphorus (P) concentrations were measured in supernatant of a water slurry (1:1, w:v, shaken for 18 h at room temperature) by malachite green method (Ohno and Zibilske, 1991). Total P concentrations were measured by the same method in acid (HNO_3 , 69%) extracts of the samples following incineration at 550°C for 20 h. Reported values for P concentrations are means of two separate extractions performed 9 and 19 days after application of the mineral P-fertilizer. The carbon (C) and nitrogen (N) concentrations in the samples were analyzed by Flash EA 2000 elemental analyzer (Thermo Fisher Scientific). All presented data are means of two technical replicates.

one of two P solutions, prepared by dissolving either 11.563 or 46.251 g of $\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$ per 1 L of distilled water. The 10 mL dose of the respective P solution or distilled water in case of P0 level was applied and mixed into the upper two thirds of the substrate filled into each pot simultaneously with the application of cations as described earlier. For P availability measured in the different substrates, please see **Table 1**.

Mycorrhizal Inoculation

Sixty pots were inoculated with the AM fungal isolate *R. irregularis* 'PH5.' The isolate is maintained in the AM fungal collection of the Department of Mycorrhizal Symbioses (Institute of Botany, Czech Academy of Sciences, Průhonice, Czech Republic) in sand–zeolite–LT soil (2:2:1, v:v:v) mixture. The AM fungal inoculum cultures, established with *Zea mays* as the initial and *Desmodium* sp. as the follow-up, long-term host plant, were 16 months old when used as inoculum source. Inspection under a stereomicroscope had confirmed very abundant intraradical and extraradical sporulation of *R. irregularis*, as well as an absence of contamination by other AM fungal morphospecies. To prepare the AM fungal inoculum, the host plants' shoots were removed and the roots were cut into ca 0.5 cm pieces and mixed back into the substrate. The material was subsequently dried at room temperature for 1 week. After thorough homogenization by mixing, the complex AM fungal inoculum (substrate+roots) was weighed into 50 g aliquots, stored temporarily in plastic bags, and then mixed into the upper two thirds of the substrate filled into each mycorrhiza-inoculated pot. This was done simultaneously with the application of cations and P (if applicable) described earlier.

To obtain an appropriate control (non-mycorrhizal, NM) treatment, a "mock" inoculum was produced in exactly the same manner as described above but using NM cultures: the same host plants were grown in the same substrate and under the same conditions as the AM fungal inoculum, but without the AM fungi. Visual inspection of the mock-inoculum cultures under a stereomicroscope confirmed the absence of AM fungal spores and/or mycelium clumps. The mock inoculum was processed and applied into the experimental pots in exactly the same manner as was the AM fungal inoculum (see above).

Plants and Rhizobia

The seeds of *M. truncatula* J5 and *M. sativa* cv. Vlasta were surface-sterilized (10% sodium hypochlorite; 10 min) and thereafter rinsed with sterilized tap water. The plants were germinated on moist filter paper in sterilized glass Petri dishes. Those seedlings with developed cotyledon leaves were transplanted into the pots, four seedlings per pot. During transplantation, the plants were inoculated with their compatible rhizobia. *M. truncatula* was inoculated with *Sinorhizobium meliloti* strain LT10, indigenous to LT soil, which was previously selected amongst several rhizobium strains isolated from the Litoměřice field site as the most beneficial rhizobium compatible with *M. truncatula* (unpublished observation). *M. sativa* was inoculated with strain 740 (Rhizobial collection, Crop Research Institute, Prague, Czech Republic), which had been recommended for *M. sativa* plants by Lenka Kabátová

(Crop Research Institute, Prague, Czech Republic, personal communication). Both of the bacterial strains were grown in TY liquid medium (Somasegaran and Hoben, 1994) on a shaker at 24°C for 3 days. The bacteria were washed with 0.5% (w/v) aqueous MgSO_4 solution and the suspension was then adjusted to the optical density of 0.7 at 600 nm (which corresponded to approximately 2×10^9 cells mL^{-1}). One mL of this suspension was applied to each planting pit of individual seedlings during planting. After 1 week, the plants were thinned to two plants per pot.

Additionally, two control pots were established, one with LT and the other with Tán substrate, both of which were added with the AM fungal inoculum and fertilized with 40 mg P per pot. These pots were then planted with an isogenic mutant TRV25 (Morandi et al., 2005) of *M. truncatula* with suppressed ability to form both mycorrhizal and rhizobial symbioses. The plants were treated with the LT10 rhizobial strain as were the other experimental pots planted with *M. truncatula*. These pots were important for estimating the amount of N taken up from the substrate by P-sufficient plants in the absence of BNF. Due to spatial limitations and low availability of seeds of the mutant plant genotype, such control treatment could not have been established for each P-supply level and in a fully replicated manner.

Plant Cultivation and ^{15}N Labeling

The experiment was begun at the end of September and conducted for 9 weeks in a heated greenhouse (where temperature did not drop below 18°C at night). Natural light was supplemented with 400 W metal halide lamps set to 14 h photoperiod such that the photosynthetically active radiation flux at plant level ranged between $370 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ at midday and a minimum of $85 \mu\text{mol} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ at dawn or dusk. The positions of the pots were fully randomized. The plants were watered with 25, 50, or 100 mL of distilled water per day (all pots received always the same amount of water that progressively increased with plant age).

The plants were regularly fertilized with N provided as NH_4NO_3 solution. N-fertilization was first applied in the 3rd week after planting (to prevent potential suppression of nodulation at early stages of plant development) and then repeated weekly, thus totaling six applications per pot. With each application, the plants were provided with 20 mg of N per pot (1.14 g of NH_4NO_3 was dissolved per 1 L of distilled water, and 50 mL of this solution were applied per pot).

To distinguish N uptake by plants via the root/mycorrhizal and the BNF pathways, the ammonium nitrate applied in the pots was enriched with $^{15}\text{NH}_4^{15}\text{NO}_3$ (>98% ^{15}N ; Cambridge Isotope Laboratories, Inc., Andover, MA, USA) to reach $\delta^{15}\text{N} = +4491\text{‰}$, corresponding to fractional abundance of ^{15}N of 0.01979 (calculated value using isotopic abundance of the unlabeled and ^{15}N -enriched ammonium nitrate and their molar ratio in the liquid fertilizer).

Harvest and Sampling

The shoots were cut at the substrate surface level, pooled per pot, dried at 65°C to constant weight and weighed to obtain shoot

dry weight (SDW). The compact root system with substrate was removed from the pot and the roots were shaken off to remove most of the substrate.

The roots were then carefully washed of the remaining substrate with water. Mycorrhizal colonization was assessed on roots sampled throughout the zone originally laying approximately in the 4–8 cm depth. The sampled roots were cut into ca 1 cm pieces, immersed into 10% KOH and then stained using the modified method of Koske and Gemma (1989). In brief, the roots were first macerated in 10% KOH (overnight at room temperature, then 50 min at 90°C), washed with tap water, neutralized in 2% lactic acid (20 min at 90°C), and stained with 0.05% Trypan blue in LG (lactic acid–glycerol–water, 1:1:1, v:v:v) for 30 min at 90°C plus overnight at room temperature. The next day, roots were washed with tap water and further stored in LG. Colonization was evaluated microscopically using an Olympus SZX12 dissecting microscope at 100× magnification and quantified according to the gridline intersection method (Giovannetti and Mosse, 1980) while observing at least 100 intersections per sample.

The remaining roots were also weighed fresh and then reweighed after drying at 65°C to constant weight. Root dry weight (RDW) of the entire root system per pot was then calculated. Plants' total dry weight (TDW) was calculated as the sum of SDW and RDW. To compare plants' growth response to inoculation in different substrate treatments, mycorrhizal growth response (MGR) of individual mycorrhizal pots was calculated from the TDW values according to the equation $MGR = (M - NM_{mean})/NM_{mean} \times 100\%$ (Gange and Ayres, 1999), where M is the TDW recorded for a given mycorrhizal pot and NM_{mean} is the mean TDW of pots in the corresponding NM treatment (i.e., the same substrate and P level).

Elemental Analyses

Prior to analyses of P and N concentrations in plant tissues, the dried samples of shoots and roots were ground to powder using a ball mill (MM200, Retsch, Haan, Germany). To determine the P concentration in plant tissues, milled samples of shoots and roots (100 mg each) were incinerated in a muffle furnace at 550°C for 12 h. The resulting ash was combined with 1 mL of concentrated (69%) HNO₃ and briefly heated to 250°C on a hot plate. The materials was then transferred to volumetric flasks through a filter paper and brought up to 50 mL with ultrapure (18 MΩ) water. Phosphorus concentration in the extracts was then measured by colorimetry at 610 nm using a Pharmacia LKB Ultrospec III spectrophotometer by the malachite green method (Ohno and Zibilske, 1991).

The N concentrations and N isotopic composition in shoots and roots were measured using a Flash EA 2000 elemental analyzer coupled with a Delta V Advantage isotope ratio mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA).

Total N and P contents were calculated from SDW and RDW data and the concentrations of the corresponding elements in shoot and root biomass, respectively. Additionally, mycorrhizal P-uptake response (MPR) and mycorrhizal N-uptake response

(MNR) were calculated from the P contents of the plants (shoots and roots combined) similarly as described above for the MGR.

Calculation of BNF Efficiency

Assuming very similar isotopic composition of aerial N₂ and total N in the potting substrates (fractional abundance of ¹⁵N in those two pools being 0.00364 and 0.00368, respectively, with the latter being the grand mean of 12 measurements of the two potting substrates amended with different levels of P before the experiment), the fraction of plant N derived from the ¹⁵N-labeled fertilizer (Ndff) was calculated separately for the shoots (Ndff_S) and the roots (Ndff_R) as follows:

$$Ndff_S (\text{mg N}) = \text{shoot N content (mg)} \times$$

$$(\frac{{}^{15}\text{N} - \text{AT}\%_S - 0.368}{1.979 - 0.368}) \quad (1)$$

$$Ndff_R (\text{mg N}) = \text{root N content (mg)} \times$$

$$(\frac{{}^{15}\text{N} - \text{AT}\%_R - 0.368}{1.979 - 0.368}), \quad (2)$$

where ¹⁵N–AT%_S and ¹⁵N–AT%_R represent isotopic composition of N in the shoots and roots expressed as ¹⁵N atom percent, respectively, and were measured by isotope ratio mass spectrometry.

From these values, efficiency of the BNF was calculated, here defined as the fraction of the plant N derived from biological N fixation (%N_{BNF}) as follows:

$$\%N_{BNF} (\% \text{ of the plant N}) =$$

$$\frac{[(\text{shoot N content} - Ndff_S) + (\text{root N content} - Ndff_R)]}{(\text{shoot N content} + \text{root N content})} \quad (3)$$

This calculation effectively ignored the contribution of seed N (likely to be very small due to the small seed size of the two experimental plant species) as well as the contribution of N contained in the components of the potting substrates. This simplification was necessary because we had not established replicated non-fixing controls for each and every combination of the potting substrate and P amendment to experimentally measure N acquisition from the differentially P-amended substrates by non-fixing plants. Inasmuch as we have experimental evidence that the substrate contribution to N acquisition by the plants is generally not very large, the simplifications described above are justifiable. Indeed, the measured contribution of substrate N to N uptake of the asymbiotic *M. truncatula* mutants supplied with the highest P level (and assuming this saturated their P demand) reached only 12.0 or 11.2 mg N for the LT and Tān substrates, respectively. This was only about 19 and 18% of their total N content when grown in the LT and Tān substrates, respectively, meaning the plants relied to a large extent for their N supply on N uptake from the liquid fertilizer (which was the only remaining N source for these plants lacking BNF as well as AM symbiosis). If extrapolated to our symbiotic experimental plants, N acquisition from the substrates would only cover about 12% of their N budget. The actual values

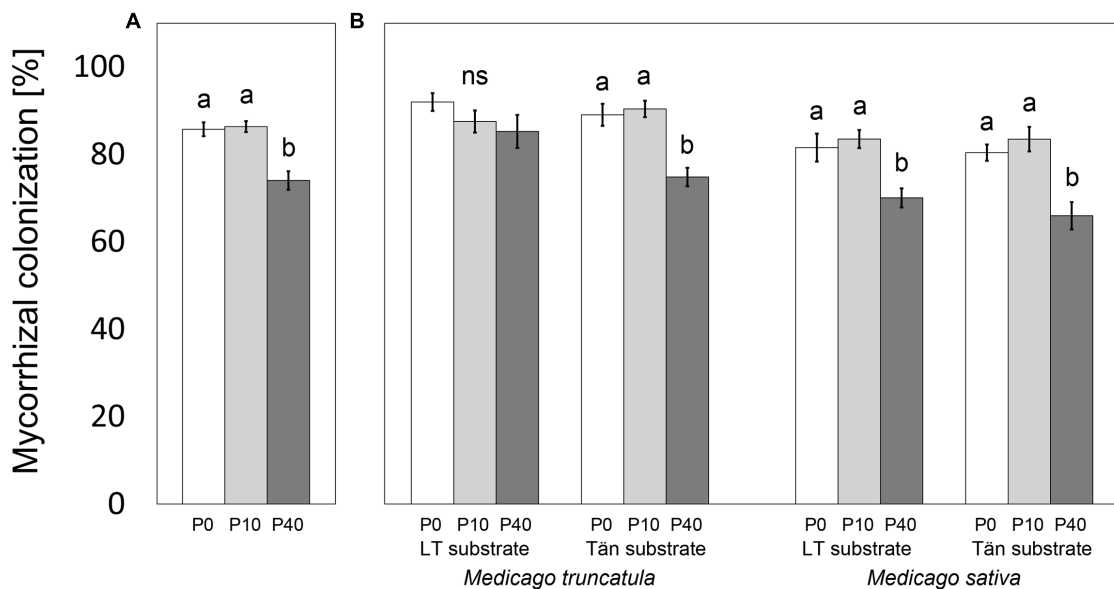


FIGURE 2 | Mycorrhizal colonization of plants inoculated with *Rhizophagus irregularis* 'PH5' affected by the addition of 0, 10, or 40 mg pot⁻¹ of phosphorus (P0, P10, and P40, respectively). The results are shown either for the whole data set (A), or separately for *Medicago truncatula* or *M. sativa* plants grown in sand–zeolite substrate with 10% of either Litoměřice or Tănikon soil (LT or Tän substrate, respectively) (B). Colonization is expressed as percentage of root length occupied by mycorrhizal fungal structures such as hyphae, arbuscules, and/or vesicles. Means \pm standard errors are shown ($n = 19$ or 20 for A, $n = 4$ or 5 for B). Different letters indicate significant differences between different P levels within each plant–substrate combination according to Tukey's HSD test ($p < 0.05$); ns, $p \geq 0.05$. The roots of control non-mycorrhizal plants remained free of mycorrhizal fungal structures (data not shown).

were most likely even lower than that 12% of the plant N budget due to the functional BNF.

Statistical Analyses

The data were analyzed using STATISTICA 12 (StatSoft Inc., USA). None of the presented data deviated significantly from the normal distribution and thus the data sets were not transformed for the statistical analyses. Data for two pots were removed from the subsequent statistical analyses: one pot from the NM treatment whose roots were colonized heavily by AM fungi (i.e., due to contamination) and one mycorrhizal pot with unusually high P content in the plant (more than twice the average of the treatment). Therefore, at least four biological replicates were retained per each treatment combination and five were included in most of them. The data were first subjected to general linear model analyses using the factors “plant,” “substrate,” and “AM fungal inoculation” as categorical predictors and “P addition” as continuous predictor in order to determine the contributions of individual factors and/or their interactions to explaining the variability in the data set (Supplementary Table S1). Individual parameters were analyzed by *t*-test to find differences between mycorrhizal and NM plants in every combination of substrate and fertilizer. A *t*-test also was used to analyze general differences in mycorrhizal colonization, MGR, MPR, and MNR between *M. truncatula* and *M. sativa* plants. The differences in mycorrhizal colonization, MGR, and MPR between P-fertilization treatments were analyzed with ANOVA followed by Tukey's HSD test for separating the means. Correlation analyses between plant P nutrition and %N_{BNF} were carried out

using a linear regression model. The slopes of regression lines for mycorrhizal and non-mycorrhizal plants were further compared in Statgraphics Plus 3.1 (Statistical Graphics Corp., USA).

RESULTS

Development of the Symbiotic Microorganisms

All plants, excluding the two pots planted with the TRV25 mutant of *M. truncatula*, had nodules well developed in their root systems. The roots of all mycorrhizal plants were also highly (>65% of the root length) colonized with AM fungi (Figure 2), whereas those of NM and the mutant plants remained free of AM fungal colonization (data not shown). Mycorrhizal *M. truncatula* plants had their roots colonized to a significantly greater extent than did the *M. sativa* plants (87% vs. 77% of root length, respectively; *t*-test, $p < 0.0001$). The levels of mycorrhizal colonization were generally significantly lower in the P40 treatment compared to the less-fertilized treatments (Figures 2A,B), the exception being for *M. truncatula* in LT substrate (Figure 2B).

Plant Growth and Mycorrhizal P Uptake

In general, the plants responded to the presence of AM fungi in terms of their biomass either positively (*M. truncatula*, except for the P40 treatment in both substrates) or else no significant effect was recorded. No case of significant negative effect of AM symbiosis on plants' TDW was found within the individual

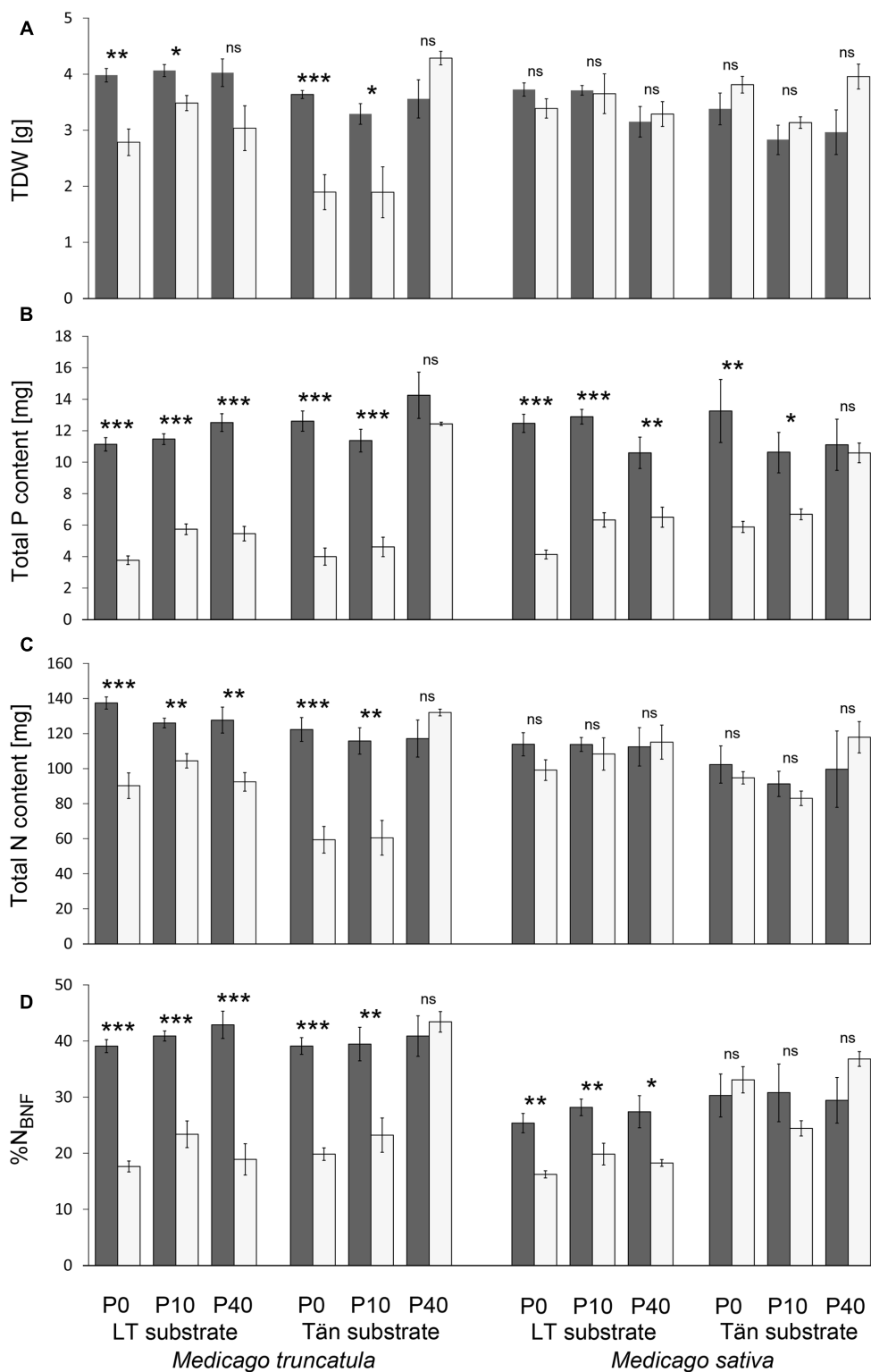


FIGURE 3 | Total plant dry weight (TDW) (A), total phosphorus (P) content (B), total nitrogen (N) content (C), and fraction of plant N derived from biological N fixation (%N_{BNF}) (D) of mycorrhizal (dark columns) or non-mycorrhizal (light columns) *M. truncatula* or *M. sativa* plants grown in sand-zeolite substrate with 10% of either Litoměřice or Tănikon soil (LT or Tăn substrate, respectively). The substrate was amended with 0, 10, or 40 mg pot⁻¹ of phosphorus (P0, P10, and P40, respectively). Bars represent means \pm standard errors ($n = 4$ or 5). Asterisks indicate significant differences between mycorrhizal and non-mycorrhizal plants according to t -test (* $0.01 \leq p < 0.05$, ** $0.001 \leq p < 0.01$, * $p < 0.001$; ns, $p \geq 0.05$).**

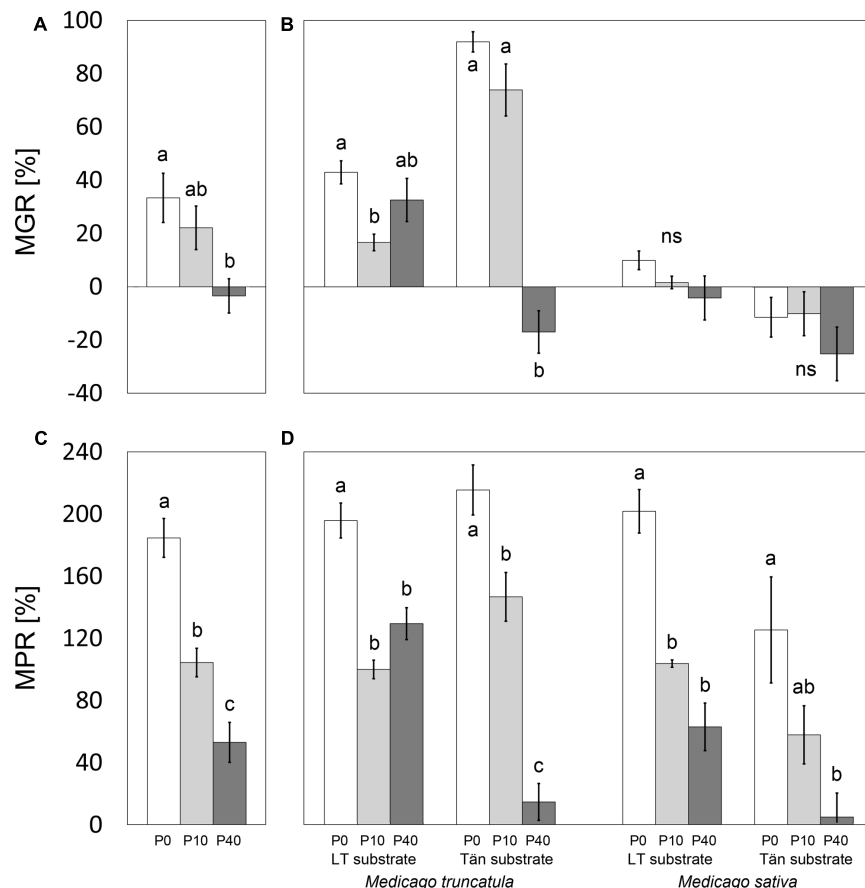


FIGURE 4 | Mycorrhizal growth response (MGR) and mycorrhizal phosphorus-uptake response (MPR) of plants along a phosphorus fertilization gradient consisting of three input levels (0, 10, and 40 mg P added per pot; P0, P10, and P40, respectively). The results are shown either for the whole data set (A,C), or separately for *M. truncatula* or *M. sativa* plants grown in sand–zeolite substrate with 10% of either Litoměřice or Tănikon soil (LT or Tăn substrate, respectively). Means \pm standard errors are shown ($n = 19$ or 20 for A and C, $n = 4$ or 5 for B and D). Different letters indicate significant differences between different P levels within each plant–substrate combination according to Tukey's HSD test ($p < 0.05$); ns, $p \geq 0.05$.

treatments (Figure 3A). The calculation of MGR nevertheless did show negative values in some cases, and particularly for the P40 treatments (Figures 4A,B). MGR of plants, analyzed for the whole data set covering four combinations of plant species and substrates, was significantly negatively correlated with the increasing P inputs ($R^2 = 0.1663$, $p = 0.013$). This general trend was driven, however, by a very strong correlation ($R^2 = 0.901$, $p < 0.0001$) recorded for *M. truncatula* planted in Tăn substrate, whereas correlations for the other plant–substrate combinations were not significant (data in Figure 4B). Highly significant differences in MGR were found between *M. truncatula* and *M. sativa* plant species, the former responding much more strongly to mycorrhiza than the latter (t -test, $p < 0.0001$).

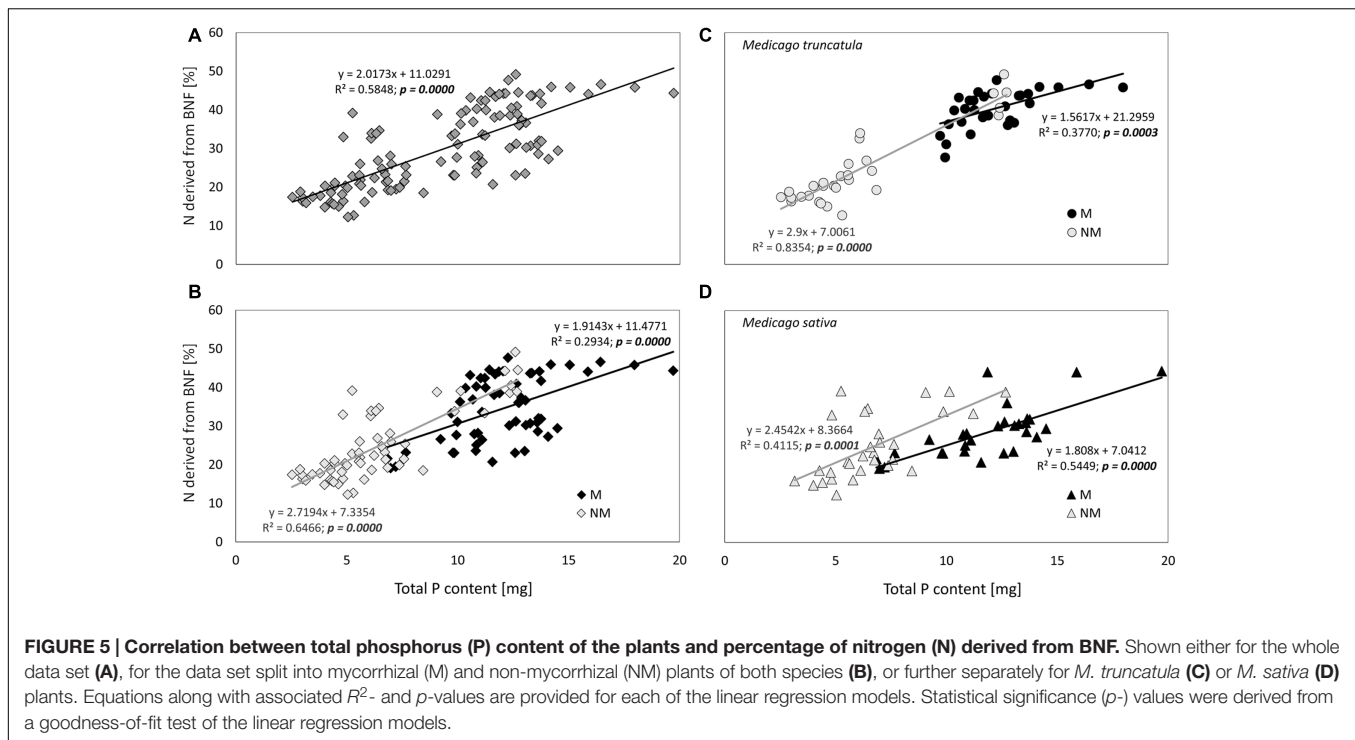
Mycorrhizal symbiosis significantly increased P uptake by both *Medicago* species in all substrate and fertilization treatments, except for the P40 treatment in Tăn substrate (Figure 3B). MPR was significantly higher in *M. truncatula* plants than in *M. sativa* plants (t -test, $p = 0.0413$). A strong negative correlation ($R^2 = 0.4566$, $p < 0.0001$) between MPR and P-fertilization was found for the whole data set (Figure 4C). Individual

plant–substrate combinations followed this trend (Figure 4D). Only in the case of *M. truncatula* planted in LT substrate was the correlation not significant (data in Figure 4D).

Biological Nitrogen Fixation

With the exception of the P40 treatment in Tăn substrate, mycorrhizal symbiosis increased total plant N content and %N_{BNF} in all *M. truncatula* plants (Figures 3C,D). In *M. sativa* plants, by contrast, the presence of AM fungus had no effect whatsoever on N content (Figure 3C), but it increased %N_{BNF} in the LT substrate irrespective of the P input level (Figure 3D). Also, highly significant differences in MNR (t -test, $p < 0.0001$) evidenced the more important role of mycorrhiza in N acquisition by *M. truncatula* plants than by *M. sativa* plants.

The %N_{BNF} was strongly and positively correlated with P content in plant biomass. This was manifest not only for the data set as a whole (Figure 5A), but it was confirmed also when smaller data sets were tested separately (Figures 5B–D). While in the case of *M. truncatula* plants the slopes of regression lines for mycorrhizal and NM plants differed significantly ($p = 0.005$),



and with the regression line for NM plants being steeper than that for mycorrhizal plants (Figure 5C), in the case of *M. sativa* plants the slopes of the regression lines for mycorrhizal and NM plants were not statistically different (Figure 5D). Likewise, the slopes of regression lines for mycorrhizal and NM plants pooled across the two plant species (Figure 5B) did not differ significantly ($p > 0.05$).

DISCUSSION

Using two different substrates and three levels of P supply allowed establishing a wide range of experimental conditions (P availabilities) for examining the symbiotic functioning of *Medicago* spp. along a P-fertilization gradient. The two *Medicago* species differed in their response to AM symbiosis, with *M. truncatula* being substantially more responsive to mycorrhiza formation than *M. sativa* in terms of growth, P acquisition, as well as N uptake. The two soils employed in this study as substrate components caused the P-sorption kinetics to differ between the two substrates (Table 1). Presumably, P was more efficiently immobilized in the calcareous LT soil with pH 7.88 (Püschel et al., 2016) than in the acidic Tăn Luvisol with pH 6.2 (Jansa et al., 2003b). If the Tăn substrate was fertilized with 40 mg kg^{-1} P, the water extractable P levels exceeded 10 mg kg^{-1} , thereby resulting in P-sufficient conditions even for the NM plants (Table 1 and Figure 3).

Are BNF and Plant P Nutrition Related?

Considering the high P demand of the symbiotic BNF (Divito and Sadras, 2014), we had hypothesized that leguminous plants better

supplied with P would, consequently, also show higher $\%N_{\text{BNF}}$. Not only was this hypothesis clearly confirmed for both *Medicago* species in association with their own compatible rhizobia, this general trend was also valid for both mycorrhizal and NM plants (Figure 5). These results confirmed previous observations (Ankomah et al., 1996; Vadez et al., 1999; Kuang et al., 2005) made with different leguminous plant species, although the range of environmental conditions (such as P availabilities) was usually more restricted in the previously published case studies than in our current research. Duplication of BNF efficiency due to massive P-fertilization in a mixed clover-grass sward was previously reported from a mesocosm experiment by Edwards et al. (2006). That study indirectly confirmed that the increase of BNF efficiency from 25 to 50% observed in our study due to removal of P limitation for the plants – either through AM symbiosis establishment or P-fertilization – is comparable to the effects observed under other (field-relevant) settings.

We also had expected to observe functional synergy between the two root symbionts (Barea et al., 2005), particularly if the plants were exposed to P deficiency. Our experimental evidence fully supports this second hypothesis for *M. truncatula* plants but only partly so for *M. sativa* plants. In the case of *M. truncatula*, mycorrhizal plants in all treatments with low P availability (i.e., below 10 mg kg^{-1}) had significantly higher $\%N_{\text{BNF}}$ than did their respective non-mycorrhizal counterparts. This was not the same, however, for *M. sativa* plants. Although AM symbiosis still provided *M. sativa* plants with more P in exactly the same combinations of substrate and fertilization as in the case of *M. truncatula* plants, this advantage was reflected in higher $\%N_{\text{BNF}}$ in LT substrate only but not in any of the P treatments

within the Tăn substrate (**Figure 3**). This indicates that different functional traits of plants, the rhizobia, or the interaction between the two can respond to the outer environment in a context-specific manner. It seems, in fact, that *M. sativa* with its rhizobia particularly liked the Tăn substrate, as it maintained BNF levels high in this substrate irrespective of the P-supply levels.

Under ample P supply, mycorrhizal benefits in terms of improved plant P acquisition were reduced or vanished completely (**Figures 3, 4**). This is consistent with the general consensus that mineral fertilization may render root symbionts dispensable (Morgan et al., 2005). Yet, the efficiency of BNF did not necessarily follow the same trend. Careful inspection of the regression lines plotted in **Figure 5** reveals that there were different slopes of regression lines describing how P content of *M. truncatula* related to the %N_{BNF} of the same plants ($p = 0.005$). A similar observation (though only marginally significant, with $p = 0.088$) was made also for the data set as a whole, but the slopes were not significantly different between mycorrhizal and non-mycorrhizal plants of *M. sativa* ($p = 0.30$). These results indicate that, at least in the case of *M. truncatula* (**Figure 5C**), to achieve the same BNF efficiency, mycorrhizal plants had to have substantially greater P content than did the NM plants, and the maximum %N_{BNF} values were achieved with greater difficulty or more slowly for the mycorrhizal as compared to the NM plants. These results indicate that with increasing P supply, the AM fungi and rhizobia increasingly competed for another limiting resource (at least in *M. truncatula* that also showed greater root colonization levels than did *M. sativa*). It is conceivable, based on the evidence of previous research, that the elusive limiting resource for BNF under sufficient P supply is the plant C (Morgan et al., 2005; Mortimer et al., 2009). If the metabolic energy to fix atmospheric N₂ is in short supply due to significant mycorrhizal C sink, which could be as high as 20% of the gross photosynthetic production (Jakobsen and Rosendahl, 1990), the benefits conferred to the host by rhizobia are actually hampered by the AM fungi. Although we do not have unequivocal evidence to show that this is happening, it is highly plausible, based also on previous experimental evidence showing additivity of C costs of the two microsymbionts in tripartite root symbioses (Paul and Kucey, 1981; Mortimer et al., 2008; Millar and Ballhorn, 2013; Ballhorn et al., 2016). Under high P availability or low light conditions, the coexistence of two root symbionts becomes a burden for the plant host (Ballhorn et al., 2016). Indirect support for this theory can be observed in the suppression of root colonization by AM fungus in most of the plant–substrate treatments with increasing P-fertilization (**Figure 2**), which is consistent with the preferential allocation hypothesis (Bever, 2015).

Relatives, Yet Functionally Different

Although using two different species of the genus *Medicago* yielded strong evidence here for a common underlying mechanism with respect to the functional interactions between mycorrhizal symbiosis and the BNF along a P-availability gradient, there were also some notable differences. One important issue that needs to be emphasized here is that both the plant and the rhizobial genotypes differed for the two plant

species treatments. This was intentionally established in this manner to achieve the highest functional compatibility of the plant–bacterial partners. Thus, we were actually comparing two plant–rhizobial (biological) systems rather than two plant species *per se*.

Non-mycorrhizal *M. sativa* plants cultivated in the Tăn substrate yielded surprisingly high %N_{BNF} despite that their P content under low P supply was significantly smaller than that of their mycorrhizal counterparts (**Figure 3**). It is possible that the rhizobia associated with *M. sativa* were either less P-demanding, more P-efficient, or generally more adapted to specific conditions of Tăn substrate than were the bacterium used to inoculate *M. truncatula*. Such differences have been described previously and have been argued to be the result of plant–bacterial coevolution (Garau et al., 2005). Interestingly, the N contents and biomass of mycorrhizal and NM *M. sativa* plants were surprisingly similar in all substrate treatments, even though the BNF efficiency and P uptake obviously varied markedly (**Figure 3**). We therefore assume that *M. sativa* might actually better compensate for the missing symbiotic benefits through more dynamic root traits such as greater plasticity of root branching (Lynch, 2007; Nibau et al., 2008) and/or root exudation (Rao et al., 2016). Inasmuch as these traits were not recorded in our study, however, this remains a matter of speculation, although it does point to possible mechanisms accounting for why different plants vary in their response and/or dependency on mycorrhizal and other symbioses (Linderman and Davis, 2004; Jakobsen et al., 2005).

CONCLUSION

Working with a large environmental (P availability) gradient established by using two different kinds of substrates in combination with three levels of mineral P inputs, we show here that AM symbiosis clearly promotes BNF efficiency, particularly in the case of low P supply. This effect was most likely mediated by improved P acquisition of the mycorrhizal as compared to the NM plants under conditions of low P. With increasing P inputs, however, the costs of the AM symbiosis (at least in the more heavily colonized *M. truncatula*) become more and more apparent, resulting in a lower P use efficiency (or in luxurious P uptake) of the mycorrhizal plants as compared to the NM plants and without concomitant increases in plant biomass production. Based upon Liebig's law of the minimum (Johnson et al., 2015), therefore, we conclude that there was strong competition between the symbionts and the plants for another resource, thereby preventing the occurrence of a significant positive growth response in the plants at higher P-supply levels. Most likely, this competition was for carbon (Ballhorn et al., 2016). In response to sufficient (or even luxurious) P supply at higher P-fertilization levels, mycorrhizal root colonization levels were reduced. Although this was in accordance with previous reports (Treseder, 2004), this obviously was not effective enough to counteract the C drain to the AM fungus, at least not in the *M. truncatula*. Although general reduction of root colonization at higher P-fertilization levels was true for both *Medicago*

species (each in association with its compatible rhizobia), notable differences were observed between the two plant species. These could be due either to the plants or the rhizobial strains used and reflect such factors as inherent tolerance of the particular rhizobia to deviation from their pH optima, different architecture of the root systems, differential efficiency of plant genotypes in mineral nutrient use and/or redistribution, root exudation patterns, or other mechanisms. Disentangling these factors would require further experimental efforts, and particularly with respect to quantifying the C costs of the two root symbioses under a range of environmental conditions.

AUTHOR CONTRIBUTIONS

DP, MJ, and JJ designed the experiment, which was then carried out mainly by DP. HG conducted the P and N analyses, and JJ calculated the BNF efficiency. DP conducted the statistical analyses. All authors contributed to interpreting the results. DP and JJ did most of the writing, whereas MJ,

AV, and MV critically commented on earlier versions of the manuscript.

ACKNOWLEDGMENTS

This research was carried out in a joint working group involving the Institute of Microbiology and the Institute of Botany. It was financially supported by the Czech Science Foundation (project 15-05466S), Czech Ministry of Education, Youth and Sports (project LK11224), and the J. E. Purkyně fellowship to JJ, as well as by the long-term research development programs RVO 61388971 and RVO 67985939.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00390/full#supplementary-material>

REFERENCES

- Ankomah, A. B., Zapata, F., Hardarson, G., and Danso, S. K. A. (1996). Yield, nodulation, and N-2 fixation by cowpea cultivars at different phosphorus levels. *Biol. Fertil. Soils* 22, 10–15. doi: 10.1007/BF00384426
- Ballhorn, D. J., Schädler, M., Elias, J. D., Millar, J. A., and Kautz, S. (2016). Friend or foe - Light availability determines the relationship between mycorrhizal fungi, rhizobia and lima bean (*Phaseolus lunatus* L.). *PLoS ONE* 11:e0154116. doi: 10.1371/journal.pone.0154116
- Barea, J. M., Pozo, M. J., Azcón, R., and Azcón-Aguilar, C. (2005). Microbial co-operation in the rhizosphere. *J. Exp. Bot.* 56, 1761–1778. doi: 10.1093/jxb/eri197
- Bethlenfalvay, G. J., Pacovsky, R. S., Bayne, H. G., and Stafford, A. E. (1982). Interactions between nitrogen-fixation, mycorrhizal colonization, and host-plant growth in the *Phaseolus*-*Rhizobium*-*Glomus* symbiosis. *Plant Physiol.* 70, 446–450. doi: 10.1104/pp.70.2.446
- Bever, J. D. (2015). Preferential allocation, physio-evolutionary feedbacks, and the stability and environmental patterns of mutualism between plants and their root symbionts. *New Phytol.* 205, 1503–1514. doi: 10.1111/nph.13239
- Bolger, T. P., Pate, J. S., Unkovich, M. J., and Turner, N. C. (1995). Estimates of seasonal nitrogen-fixation of annual subterranean clover-based pastures using the 15N natural-abundance technique. *Plant Soil* 175, 57–66. doi: 10.1007/BF02413010
- Divito, G. A., and Sadras, V. O. (2014). How do phosphorus, potassium and sulphur affect plant growth and biological nitrogen fixation in crop and pasture legumes? A meta-analysis. *Field Crop Res.* 156, 161–171. doi: 10.1016/j.fcr.2013.11.004
- Edwards, E. J., McCaffery, S., and Evans, J. R. (2006). Phosphorus availability and elevated CO2 affect biological nitrogen fixation and nutrient fluxes in a clover-dominated sward. *New Phytol.* 169, 157–167. doi: 10.1111/j.1469-8137.2005.01568.x
- Gange, A. C., and Ayres, R. L. (1999). On the relation between arbuscular mycorrhizal colonization and plant 'benefit'. *Oikos* 87, 615–621. doi: 10.2307/3546829
- Garau, G., Reeve, W. G., Brau, L., Deiana, P., Yates, R. J., James, D., et al. (2005). The symbiotic requirements of different *Medicago* spp. suggest the evolution of *Sinorhizobium meliloti* and *S. medicae* with hosts differentially adapted to soil pH. *Plant Soil* 276, 263–277. doi: 10.1007/s11104-005-0374-0
- Giovannetti, M., and Mosse, B. (1980). Evaluation of techniques for measuring vesicular arbuscular mycorrhizal infection in roots. *New Phytol.* 84, 489–500. doi: 10.1111/j.1469-8137.1980.tb04556.x
- Jakobsen, I. (1985). The role of phosphorus in nitrogen fixation by young pea plants (*Pisum sativum*). *Physiol. Plant.* 64, 190–196. doi: 10.1111/j.1399-3054.1985.tb02334.x
- Jakobsen, I., Chen, B. D., Munkvold, L., Lundsgaard, T., and Zhu, Y. G. (2005). Contrasting phosphate acquisition of mycorrhizal fungi with that of root hairs using the root hairless barley mutant. *Plant Cell Environ.* 28, 928–938. doi: 10.1111/j.1365-3040.2005.01345.x
- Jakobsen, I., and Rosendahl, L. (1990). Carbon flow into soil and external hyphae from roots of mycorrhizal cucumber plants. *New Phytol.* 115, 77–83. doi: 10.1111/j.1469-8137.1990.tb00924.x
- Jansa, J., Finlay, R., Wallander, H., Smith, F. A., and Smith, S. E. (2011). Role of mycorrhizal symbioses in phosphorus cycling. *Soil Biol.* 26, 137–168. doi: 10.1007/978-3-642-15271-9_6
- Jansa, J., Mozafar, A., and Frossard, E. (2003a). Long-distance transport of P and Zn through the hyphae of an arbuscular mycorrhizal fungus in symbiosis with maize. *Agronomie* 23, 481–488. doi: 10.1051/agro:2003013
- Jansa, J., Mozafar, A., Kuhn, G., Anken, T., Ruh, R., Sanders, I. R., et al. (2003b). Soil tillage affects the community structure of mycorrhizal fungi in maize roots. *Ecol. Appl.* 13, 1164–1176. doi: 10.1007/s12275-015-5108-2
- Johnson, N. C., Graham, J. H., and Smith, F. A. (1997). Functioning of mycorrhizal associations along the mutualism-parasitism continuum. *New Phytol.* 135, 575–586. doi: 10.1046/j.1469-8137.1997.00729.x
- Johnson, N. C., Wilson, G. W. T., Wilson, J. A., Miller, R. M., and Bowker, M. A. (2015). Mycorrhizal phenotypes and the Law of the Minimum. *New Phytol.* 205, 1473–1484. doi: 10.1111/nph.13172
- Kaschuk, G., Kuyper, T. W., Leffelaar, P. A., Hungria, M., and Giller, K. E. (2009). Are the rates of photosynthesis stimulated by the carbon sink strength of rhizobial and arbuscular mycorrhizal symbioses? *Soil Biol. Biochem.* 41, 1233–1244. doi: 10.1016/j.soilbio.2009.03.005
- Kaschuk, G., Leffelaar, P. A., Giller, K. E., Alberton, O., Hungria, M., and Kuyper, T. W. (2010). Responses of legumes to rhizobia and arbuscular mycorrhizal fungi: a meta-analysis of potential photosynthate limitation of symbioses. *Soil Biol. Biochem.* 42, 125–127. doi: 10.1016/j.soilbio.2009.10.017
- Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., et al. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* 333, 880–882. doi: 10.1126/science.1208473
- Kleinert, A., Venter, M., Kossmann, J., and Valentine, A. (2014). The reallocation of carbon in P deficient lupins affects biological nitrogen fixation. *J. Plant Physiol.* 171, 1619–1624. doi: 10.1016/j.jplph.2014.07.017

- Koske, R. E., and Gemma, J. N. (1989). A modified procedure for staining roots to detect VA-mycorrhizas. *Mycol. Res.* 92, 486–505. doi: 10.1016/S0953-7562(89)80195-9
- Kuang, R. B., Liao, H., Yan, X. L., and Dong, Y. S. (2005). Phosphorus and nitrogen interactions in field-grown soybean as related to genetic attributes of root morphological and nodular traits. *J. Integr. Plant Biol.* 47, 549–559. doi: 10.1111/j.1744-7909.2005.00072.x
- Larimer, A. L., Bever, J. D., and Clay, K. (2010). The interactive effects of plant microbial symbionts: a review and meta-analysis. *Symbiosis* 51, 139–148. doi: 10.1007/s13199-010-0083-1
- Larimer, A. L., Clay, K., and Bever, J. D. (2014). Synergism and context dependency of interactions between arbuscular mycorrhizal fungi and rhizobia with a prairie legume. *Ecology* 95, 1045–1054. doi: 10.1890/13-0025.1
- Linderman, R. G., and Davis, E. A. (2004). Varied response of marigold (*Tagetes* spp.) genotypes to inoculation with different arbuscular mycorrhizal fungi. *Sci. Hortic.* 99, 67–78. doi: 10.1016/S0304-4238(03)00081-5
- Lynch, J. P. (2007). Roots of the second green revolution. *Aust. J. Bot.* 55, 493–512. doi: 10.1071/BT06118
- Millar, J. A., and Ballhorn, D. J. (2013). Effect of mycorrhizal colonization and light limitation on growth and reproduction of lima bean (*Phaseolus lunatus* L.). *J. Appl. Bot. Food Qual.* 86, 172–179.
- Morandi, D., Prado, E., Sagan, M., and Duc, G. (2005). Characterisation of new symbiotic *Medicago truncatula* (Gaertn.) mutants, and phenotypic or genotypic complementary information on previously described mutants. *Mycorrhiza* 15, 283–289. doi: 10.1007/s00572-004-0331-4
- Morgan, J. A. W., Bending, G. D., and White, P. J. (2005). Biological costs and benefits to plant-microbe interactions in the rhizosphere. *J. Exp. Bot.* 56, 1729–1739. doi: 10.1093/jxb/eri205
- Mortimer, P. E., Perez-Fernandez, M. A., and Valentine, A. J. (2008). The role of arbuscular mycorrhizal colonization in the carbon and nutrient economy of the tripartite symbiosis with nodulated *Phaseolus vulgaris*. *Soil Biol. Biochem.* 40, 1019–1027. doi: 10.1016/j.soilbio.2007.11.014
- Mortimer, P. E., Pérez-Fernández, M. A., and Valentine, A. J. (2009). Arbuscular mycorrhizae affect the N and C economy of nodulated *Phaseolus vulgaris* (L.) during NH₄⁺ nutrition. *Soil Biol. Biochem.* 41, 2115–2121. doi: 10.1016/j.soilbio.2009.07.021
- Nibau, C., Gibbs, D. J., and Coates, J. C. (2008). Branching out in new directions: the control of root architecture by lateral root formation. *New Phytol.* 179, 595–614. doi: 10.1111/j.1469-8137.2008.02472.x
- Ohno, T., and Zibilske, L. M. (1991). Determination of low concentrations of phosphorus in soil extracts using malachite green. *Soil Sci. Soc. Am. J.* 55, 892–895. doi: 10.2136/sssaj1991.03615995005500030046x
- Paul, E. A., and Kucey, R. M. N. (1981). Carbon flow in plant microbial associations. *Science* 213, 473–474. doi: 10.1126/science.213.4506.473
- Püschel, D., Janoušková, M., Hujslová, M., Slavíková, R., Gryndlerová, H., and Jansa, J. (2016). Plant–fungus competition for nitrogen erases mycorrhizal growth benefits of *Andropogon gerardii* under limited nitrogen supply. *Ecol. Evol.* 6, 4332–4346. doi: 10.1002/ece3.2207
- Rao, I. M., Miles, J. W., Beebe, S. E., and Horst, W. J. (2016). Root adaptations to soils with low fertility and aluminium toxicity. *Ann. Bot.* 118, 593–605. doi: 10.1093/aob/mcw073
- Raven, J. A., and Edwards, D. (2001). Roots: evolutionary origins and biogeochemical significance. *J. Exp. Bot.* 52, 381–401. doi: 10.1093/jexbot/52.suppl_1.381
- Reinhard, S., Martin, P., and Marschner, H. (1993). Interactions in the tripartite symbiosis of pea (*Pisum sativum* L.), Glomus and Rhizobium under nonlimiting phosphorus supply. *J. Plant Physiol.* 141, 7–11. doi: 10.1016/S0176-1617(11)80844-8
- Řezáčová, V., Konvalinková, T., and Jansa, J. (2017). “Carbon fluxes in mycorrhizal plants,” in *Mycorrhiza*, eds A. Varma, R. Prasad, and N. Tuteja (Cham: Springer International Publishing).
- Saia, S., Amato, G., Frenda, A. S., Giambalvo, D., and Ruisi, P. (2014). Influence of arbuscular mycorrhizae on biomass production and nitrogen fixation of berseem clover plants subjected to water stress. *PLoS ONE* 9:e90738. doi: 10.1371/journal.pone.0090738
- Slavíková, R., Püschel, D., Janoušková, M., Hujslová, M., Konvalinková, T., Gryndlerová, H., et al. (2016). Monitoring CO₂ emissions to gain a dynamic view of carbon allocation to arbuscular mycorrhizal fungi. *Mycorrhiza* 27, 35–51. doi: 10.1007/s00572-016-0731-2
- Smith, S. E., and Read, D. J. (2008). *Mycorrhizal Symbiosis*. Cambridge: Academic Press.
- Somasegaran, P., and Hoben, H. J. (1994). *Handbook for Rhizobia: Methods in Legume-Rhizobium Technology*. New York, NY: Springer. doi: 10.1007/978-1-4613-8375-8
- Treseder, K. K. (2004). A meta-analysis of mycorrhizal responses to nitrogen, phosphorus, and atmospheric CO₂ in field studies. *New Phytol.* 164, 347–355. doi: 10.1111/j.1469-8137.2004.01159.x
- Vadez, V., Lasso, J. H., Beck, D. P., and Drevon, J. J. (1999). Variability of N₂-fixation in common bean (*Phaseolus vulgaris* L.) under P deficiency is related to P use efficiency. *Euphytica* 106, 231–242. doi: 10.1023/A:1003512519558
- van der Heijden, M. G. A., de Bruin, S., Luckerhoff, L., van Logtestijn, R. S. P., and Schlaeppli, K. (2016). A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. *ISME J.* 10, 389–399. doi: 10.1038/ismej.2015.120

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Püschel, Janoušková, Voříšková, Gryndlerová, Vosátka and Jansa. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genotypic Differences in Phosphorus Efficiency and the Performance of Physiological Characteristics in Response to Low Phosphorus Stress of Soybean in Southwest of China

OPEN ACCESS

Tao Zhou^{1†}, Yongli Du^{1†}, Shoaib Ahmed^{1†}, Ting Liu¹, Menglu Ren¹, Weiguo Liu^{1*} and Wenyu Yang^{1,2*}

Edited by:
Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:
Bingcheng Xu,
Institute of Soil and Water
Conservation (CAS), China
Amanullah,
University of Agriculture, Peshawar,
Pakistan

***Correspondence:**
Weiguo Liu
lwgsy@126.com
Wenyu Yang
wenyu.yang@263.net

[†]These authors have contributed
equally to this work.

Specialty section:
This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 12 September 2016
Accepted: 10 November 2016
Published: 24 November 2016

Citation:
Zhou T, Du Y, Ahmed S, Liu T, Ren M,
Liu W and Yang W (2016) Genotypic
Differences in Phosphorus Efficiency
and the Performance of Physiological
Characteristics in Response to Low
Phosphorus Stress of Soybean in
Southwest of China.
Front. Plant Sci. 7:1776.
doi: 10.3389/fpls.2016.01776

¹ College of Agronomy, Sichuan Agricultural University, Chengdu, China, ² Key Laboratory of Crop Ecophysiology and Farming System in Southwest, Ministry of Agriculture, Chengdu, China

Southwest of China is one of the major soybean (*Glycine max* L.) production regions in China with low availability of soil phosphorus (P). Whereas little information is available on P-efficient soybean genotypes in this region, even though using P-efficient soybean genotypes is a sustainable P management strategy for enhancing yield and P use efficiency. To assess the genetic variation on P use efficiency, 274 soybean genotypes were employed to compare the yields and P acquisition potentials in the field. Additionally, 10 representational genotypes (5 P-efficient genotypes and 5 P-inefficient genotypes) were grown in hydroponic media containing low P treatment (0.05 mM L⁻¹) and high P treatment (0.25 mM L⁻¹) to further investigate P assimilation characteristics and the related mechanisms of P-efficient soybean genotypes. In the field trial, the models described the relationships between yield and seed P concentration ($R^2 = 0.85$), shoot P accumulation ($R^2 = 0.84$), HI ($R^2 = 0.82$) well. The yield, seed P concentration and shoot P accumulation ranged from 5.5 to 36.0 g plant⁻¹, from 0.045 to 0.93% and from 0.065 to 0.278 mg plant⁻¹, respectively. In the hydroponic trial, P-efficient genotypes under low P treatment showed significantly better plant growth, P accumulation and root: shoot ratio than P-inefficient genotypes. Simultaneously, total root length, specific root length, root surface area and root volume of P-efficient were significantly greater than P-inefficient under low P treatment. Higher rate of organic acid exudation and acid phosphatase activities was observed in the P-efficient soybean genotypes under low P condition when compared to the P-inefficient soybean genotypes. It indicated that significant genetic variation for P use efficiency existed in this region, and the P-efficient soybean genotypes, especially E311 and E141, demonstrated great tolerance to P deficiency, which could be potential materials using in improving production and P use efficiency in low availability of soil P region.

Keywords: *Glycine max*, phosphorus, yield, root morphology, organic acids, acid phosphatase

INTRODUCTION

Low phosphorus (P) availability existed in many soils as inherent P deficiency and/ or P strongly bound to soil particles. Crop production generally relies on regular application of P fertilizer, as P is one of the essential nutrients for crop growth. During the period of 1960–2008, the total grain output of China increased 4.4-fold from 110 to 483 million tons (FAO, 2011), which paid a price of 91-fold P increase (Zhang et al., 2012). Generally, only 15–20% of P can be taken up by crops in the season of application (Zhang et al., 2008). The remaining P is only partially available to crops, while most of them are bound to Al/Fe oxides in acidic soils or to Ca carbonate in alkaline soil and accumulate in soil (Li et al., 2011). Thus, although soil total P concentration ranges from 0.2 g P kg⁻¹ to 3.0 g P kg⁻¹, less than 1% of the total P is available for crop growth (Otani and Ae, 1996). In fact, P fertilizer is derived from mined phosphate rock, which is a finite resource and is slowly depleting (Cordell et al., 2009). Based on this, Chinese government recently encourages the farmers to decrease chemical fertilizer input in soil, conversely excavating soil P biological availability for improving P use efficiency. More and more attentions were paid on the efficient utilization of P resources (Cordell et al., 2009).

Southwest of China located at the upstream of the Yangtze River and almost the arable field on the hilly landscape with the slow economic development. Besides less P input, tremendous soil erosion by intensive cultivation and rainfall result in soil P deficiency in this region (Zhang et al., 2004; Lin et al., 2009). In 1980, the national average for soil Olsen-P was 7.4 mg kg⁻¹. In 2006, the increase of soil Olsen P occurred in all agroecological regions of China. The soil Olsen P in the middle-lower Yangtze plain, in the North China Plain and in the South China was 17.5, 20.7, and 25.4 mg kg⁻¹, respectively (Li et al., 2011). However, in 2012, the average of soil Olsen P in Sichuan province was only 16.0 mg kg⁻¹. Therefore, excavating soil P biological availability for improving productivity and P use efficiency is vital in the Southwest of China as high risk of P leaching and low P input existed in this area.

Utilization of P-efficient crops has been proved as an effective way to improve P use efficiency. Nutrient-efficient plants are defined as plants which could produce higher yields per nutrient applied or absorbed more compared with other plants grown in similar agroecological conditions (Fageria et al., 2008). Plant root plays the dominant role in increasing in soil P bioavailability and plant P acquisition, and the physiological traits of root can adapt in response to P deficiency. A fine root system includes high length, volume, biomass, specific root length, which is benefit for exploring greater P availability by occupying huge soil volume (Wang et al., 2010). Root exudates, like organic acids and phosphatase enzymes, may also enhance P acquisition by plants. Exudation of organic acids into the rhizosphere successfully increased P availability by mobilizing conservative mineral P and organic P (Po) source, and thus improved plant P acquisition (Dinkelaker et al., 1989). Another typical response of P-efficient genotypes to P deficiency is increasing phosphatase or phytase exudation to mineralize Po (Hayes et al., 2000). Following

the development of technology, genetic engineering was used in changing plant physiological and biochemical parameters to improve P use efficiency. Over-expression of organic acids and phytase or acid phosphatase genes lead to significantly increasing in exudation from roots, and therefore enhance plant P acquisition (Koyama et al., 2000; Delhaize et al., 2009; Wasaki et al., 2009). However, most of these results lacked the application in practical production, and improving P use efficiency through genetic improvement is difficult in short time. So the screening of P-efficient plants in the natural conditions is trustworthy and efficacious pathway for pursuing high grain yield and alleviating the conflicts between the depleting P resource and food demand.

Soybean (*Glycine max* L.) not only is an essential source of protein, oil and micronutrients in human and animal diets, but also possesses a pivotal ecological function in cropping system. For instance, improving soil P availability (Xia et al., 2013), nitrogen fixation (Salvagiotti et al., 2008), soil carbon sequestration (Cong et al., 2015) and decreasing soil disease (Gao et al., 2014) for themselves and neighboring. Soybean is widely cultivated all over the world and its cultivation history in China at least 3000 years (Hymowitz, 1970). P-efficient soybean genotypes had been found in South and Northeast of China, and produced a high yield in low P condition (Zhao et al., 2004; Pan et al., 2008). However, soybean is a narrow-adaptation crop. Little information was found about P-efficient soybean genotypes in southwest of China, even though where is one of the four major soybean production region in China. Maize/soybean relay strip intercropping system is a dominant cropping system in southwest of China, which makes a weighty contribution in improving crop productions and resources utilization efficiency (Liu et al., 2015). However, soybean in this system undergoes serious shading stress by maize during the common growth stage, which results decreases in yield compared to sole soybean (Yang et al., 2014). P-efficient crops showed more tolerant to low light and P deficiency condition (Wissuwa et al., 2005). Therefore, P-efficient soybean genotypes would be potential materials in improving production and P use efficiency in low light and/ or low P availability agricultural region.

Modern varieties were selected under high input conditions on breeding stations, which may not have the capable of high nutrient use efficiency, because genes controlling traits of benefit under lower soil fertility were lost as they conveyed no advantage under very high soil fertility (Wissuwa et al., 2009). Traditional genotypes showed higher P use efficiency compared with modern varieties when grown in an unfertilized and highly P-fixing soil (Wissuwa and Ae, 2001). In present study, 274 soybean genotypes were collected from Southwest of China with low availability of soil P (Li et al., 2011), and most of these soybean genotypes are traditionally. The objectives of this study were to screening P-efficient soybean genotypes and ascertaining the factors leading to the differences between P-efficient and P-inefficient soybean genotypes. Thus, yield, characteristics of P accumulation, root morphology, activities of Apase and organic acid exudation rate of the P-efficient and P-inefficient soybean genotypes were investigated.

MATERIALS AND METHODS

Plant Material

Two hundred and seventy-four soybean genotypes (*Glycine max* L.) (Table S1) were collected from Sichuan Province, Yunnan Province, Guizhou Province, Hubei Province, Chongqing City, and Guangxi Zhuang Autonomous Region in the southwest of China (Figure 1). Most of these soybeans are traditional genotype and the gene pool was not domesticated by artificial direct or indirect purpose.

Field Experiment

The field experiment was conducted in 2015 at the experimental station of Sichuan Agricultural University in Ya'an, Sichuan Province, China (29°58'N, 102°58'E) with an altitude of 600 m above sea level. Annual mean temperature is 15.4°C with a maximum and minimum temperature of 25.4° and 6.1°C, respectively. The frost-free period is 294 days, annual precipitation is 1500 mm and potential evaporation is 838 mm. Annual sunshine is 1019 h and total solar radiation averages is 3750 MJ m⁻² year⁻¹. The experimental soil is classified as Purple soil (Luvic Xerosols).

The 274 soybean genotypes were planted in a randomized block design with 3 replicates. Each block consisted of 274 plots with the dimension of 1.5 × 4.0 m². Every plot consisted of three plant rows, and the rows spacing is 50 cm. Density of soybean was about 11 plants m⁻². Soybean was sown on early June and harvested on late October in 2015. In the last season, maize had been sown in this field, and the soil properties at the start of this study were pH (water) 6.4, organic matter concentration 30.1 g kg⁻¹, total N 1.80 g kg⁻¹, available N 110 mg kg⁻¹, Olsen-P 17.4 mg kg⁻¹, exchangeable K 91 mg kg⁻¹, and Cation Exchange Capacity 22.0 cmol kg⁻¹ of dry soil in the top 20 cm soil layer. During the growth period, all the plots were well irrigated and weeded manually with no fertilizer input.

Shoot dry matter (DM) of soybean was measured at maturity. 10 plants of soybean were sampled from the middle row of each plot. All samples were dried in an oven at 70°C and then ground for further chemical measurements. The samples were wet-digested with concentrated H₂SO₄ and H₂O₂ (30%) for P determination following the vanadomolybdate method (Page, 1982). Shoot P uptake was calculated as P concentration multiplying with shoot biomass. The harvest index (HI) was calculated as dividing the yield by the shoot biomass. Grain yield of soybean came from harvesting the remaining parts of the plot after shoot sampling.

The linear-plateau model was used to analyze the relationship between yield and HI. The linear-plateau model is defined by Equations (1) and (2) as:

$$y = a + bx \quad \text{if}(x) < c \quad (1)$$

$$y = Yp \quad \text{if}(x) \geq c \quad (2)$$

where y is HI; a is the intercept parameter; b is the slope parameter; x is the yield (g/plant); c is the critical yield, which is the interception point of the two linear segments; and Yp is the plateau value which is often 90% of the maximum HI. Equation

(1) can be interpreted as the region during which the HI responds to yield increasing and Equation (2) to the plateau region where increase of yield does not come from the increase of HI.

Hydroponic P Efficient Assays

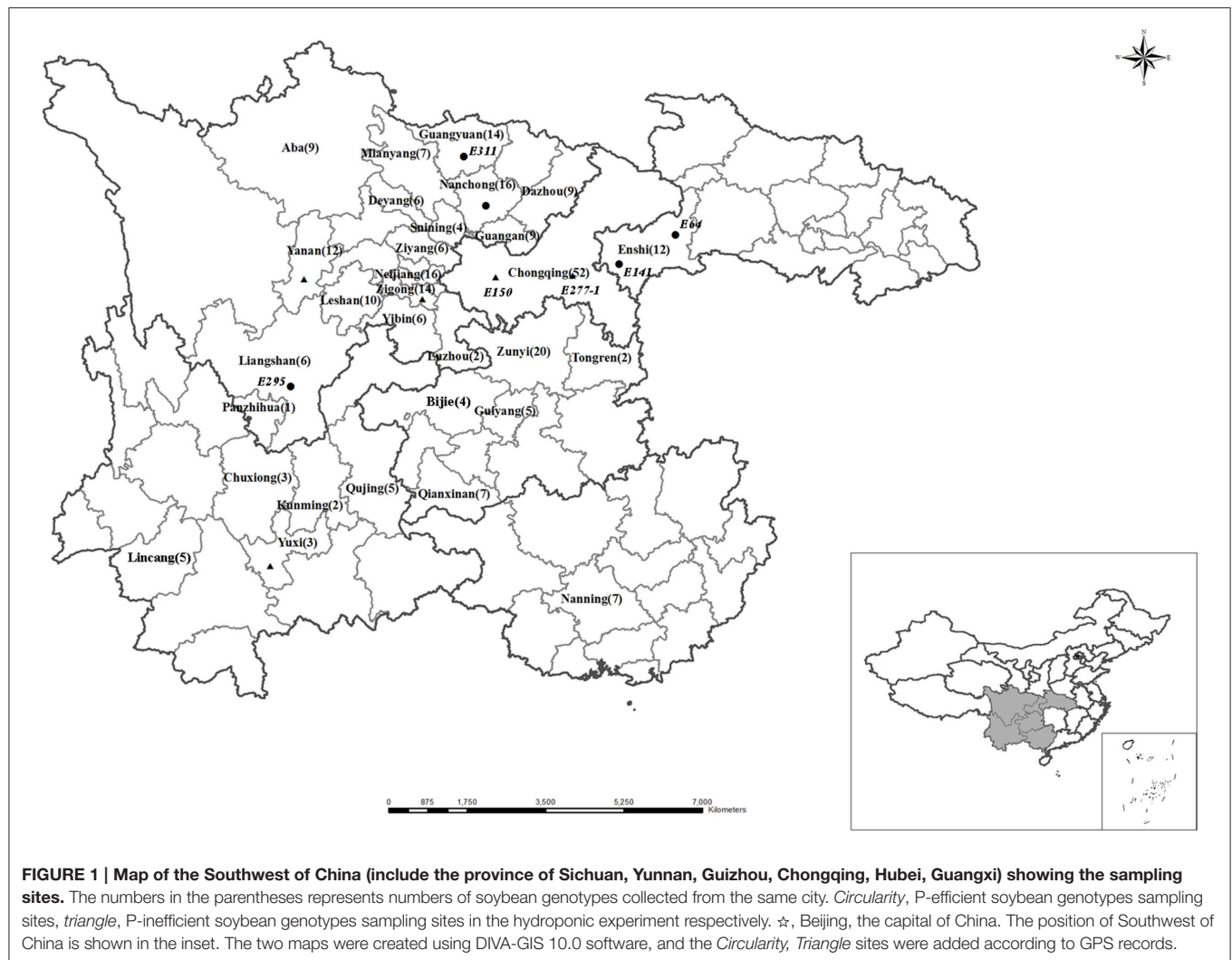
Based on yield and P accumulation of soybean genotypes in the field trial (Table S2), E14, E64, E141, E295, E311 and D55, E108, E150, E277-1, E283 representing the P-efficient, P-inefficient genotypes were used to evaluate tolerance of P deficiency in hydroponic experiment, respectively. Seeds were surface-sterilized as described by Vincent (1970), placed on sterile Whatman filter paper, and germinated in sterile water in pot filled with quartz sand until taproots were 3–4 cm long. The seedlings were transferred to a 24 liter container (0.58 × 0.28 × 0.15 m³) filled with half-strength modified Hoagland solution (0.75 mM K₂SO₄, 2 mM Ca (NO₃)₂, 0.65 mM MgSO₄, 0.1 mM KCl, 0.2 mM KH₂PO₄, 0.1 mM Fe(III)NaEDTA, 10 uM H₃BO₃, 1 uM MnSO₄, 0.1 uM CuSO₄, 1 uM ZnSO₄, 0.09 uM (NH₄)₆Mo₇O₂₄), each container with 25 soybean plants. The experiment was designed as a complete randomized with two P treatment levels (0.05 and 0.25 mM P L⁻¹) and six replicates. Plants were grown in a greenhouse from 14th February to 16th March in 2016 with an average temperature of 25/20°C (day/night), relative humidity 75%, average daytime photosynthetically active radiation between 800 and 1000 mmol m⁻² s⁻¹ and photoperiod of 14 h day/10 h night. The solution was well aerated and renewed every 5 days and pH maintained at 5.4–5.5 with daily regulation. The plants were harvested at stem elongation stage (30 days after sowing) for testing the physiological and root morphology index below. Plants show higher root secretion rate at stem elongation stage than later in growth period, when the amount of fixed C allocated to roots and rhizosphere (Gregory and Atwell, 1991).

Organic Acids in Root Exudates

For root exudation collection, three replicates of each genotype were washed absolutely with running tap water followed by distilled water and then placed separately in 100 ml glass test tube filled with deionized water and covered with black plastic to prevent light degradation of exudations. The exudate was collected for 6 h, and the details before high performance liquid chromatography (HPLC) analysis refers the method of Dong et al. (2004). Organic acids were analyzed by HPLC (Agilent 1100, Agilent, USA) after Libert and Franceschi (1987) with modifications (Yu et al., 2002). A Hypsil (Hypsil, Dalian, China) C₁₈ column (5 uM, 4.6 × 250 mm) was used as the static phase and the mobile phase was a solution containing 0.5% KH₂PO₄ and 0.5 mM tetrabutylammonium hydrogen sulfate (TBA) buffered at pH 2.0 with orthophosphoric acid. The flow rate was 1 ml min⁻¹ and detection wavelength was 220 nm.

Determination of Biomass and Root Morphology

After exudate collection, soybean plants immediately divided into shoots and roots. Shoots were dried in an oven at 70°C until constant weight. Roots were placed in clear plastic bags filled with 50% ethanol and stored in a refrigerator. Washed roots were gently arranged with minimum overlap using tweezers



in a plexiglass tray full with water, and scanned using an Epson perfection V700 photo, Japan. Images were analyzed using WinRhizo (Version 2007d, Regent Instrument Inc., Canada) to estimate total root length, root surface area, and root volume. Several images of one root were analyzed in more detail to determine the root morphology. Roots were dried in an oven at 70°C until constant weight after imaging. Specific root (cm g^{-1} DW) length was referring to the method of Pang et al. (2010).

Analysis of Root APase Activity

After harvest, fresh roots of three replicates were washed absolutely with running tap water and distilled water, and froze immediately in liquid nitrogen, then stored at -80°C . 0.3 g root tissues were ground with a mortar and pestled with 5 mL of 15 mM 2-morpholinoethanesulfonic acid, monohydrate (MES) buffer (pH 5.5, 0.5 mM $\text{CaCl}_2 \cdot \text{H}_2\text{O}$, 1 mM EDTA). The extracts were centrifuged at 4°C for 20 min at 10,000 rpm to obtain the supernatants which were used for the determination of enzyme activity.

P-nitrophenyl phosphate disodium salt hexahydrate (pNPP) was used as substrate to determinate APase activity. At first, 0.5 ml enzyme extract with a total volume of 4 ml containing 15 mM MES buffer and 10 mM pNPP was incubated at 37°C for 30 min. Subsequently, an equal volume of 0.25 M NaOH was added to terminate reaction immediately. APase activity was measured from the release of p-nitrophenol (pNP) and expressed as $\text{pNP } \mu\text{g g}^{-1}$ fresh weight (FW) min^{-1} , and pNP was determined spectrophotometrically using a UV spectrophotometer (model UN-2600A, UNICO) at 412 nm relatively to standard solutions (Sharma and Sahi, 2005).

Analysis of Tissue P Concentration

P concentration of root and shoot was analyzed following the vanadomolybdate method (Page, 1982). Shoot and root P accumulation was calculated by multiplying P concentration with the DW, respectively.

The ratio of root P accumulation: shoot P accumulation is a typical index for plant response to P deficiency. Generally, plant response to P deficiency is the increase in root: shoot ratio, which

might be due to preferential assimilate P distribution to the roots (Vance et al., 2003).

Statistical Analysis

Data from 3 replicates were sorted out by Excel (Microsoft) software packages. Regression equations were developed for the relationships between yield and seeds P concentration, shoot P uptake, HI. The liner-plateau model was analyzed by the SAS 9.1.3 software (SAS Institute Inc., USA). The liner model was analyzed by the SPSS 19.0 software. Significant difference of shoot, root dry matter and P accumulation, root length, root surface area, root volume, specific root length, activities of APase and organic acid exudation rate between soybean genotypes and P treatments were analyzed by analysis of variance (ANOVA) and mean values were compared by least significant difference (LSD) multiple comparison using the SPSS19.0 software (SPSS Institute Inc., USA).

RESULTS

Field Study

Grain Yield, Dry Matter Accumulation and P Acquisition Variable in Soybean Genotypes

Significant variation existed in grain yield among the 274 soybean genotypes in field experiment (Table S2). Average yield of 274 soybean genotypes ranging from 5.6 to 36.0 g plant⁻¹ was 16.5 g plant⁻¹. Seed P concentration and shoot P accumulation showed significant response to yield. Moreover, the linear-plateau model described the relationship between yield and HI well ($R^2 = 0.82$) (Figure 2).

To further compare the partitions of seed P concentration, shoot P accumulation, HI among different soybean varieties, pooled grain yields were divided into two yield categories: <20 g plant⁻¹ (low yield, LY) (the number of soybean genotypes, $n = 206$, mean yield: 13.6 g plant⁻¹), ≥ 20 g plant⁻¹ (high yield, HY) (the number of soybean genotypes, $n = 68$, mean yield: 25.3 g plant⁻¹). 20 g plant⁻¹ is the yield of *Nandou 12*, which is Sichuan summer soybean cultivar with the largest acreage, and the line is also the standard for choose HY soybean genotypes. Pooled seed P concentration, shoot P uptake and HI were divided into two groups below or above the mean line (Figure 2). We delimit the P-efficient soybean genotypes with high yield, shoot P accumulation, HI and seed P concentration simultaneously. On the contrary, the P-inefficient genotypes may be with low yield, shoot P accumulation, HI and seed P concentration. There are 27, 66, and 39 soybean genotypes with high (above the mean line) seed P concentration, shoot P accumulation and HI in HY category, respectively (Figure 2). However, only 16 genotypes possessed high yield, seed P concentration, shoot P accumulation and HI simultaneously. Otherwise, 99, 161, and 171 soybean genotypes possessed low (below the mean line) seed P concentration, shoot P accumulation and HI in LY category, respectively. Considering some genotypes from same site may be with same genetic background, so represent (P-efficient, P-inefficient) soybean genotypes with a large distance scale between each other were chosen. E14, E64, E141, E295, and E311 were chosen to represent the P-efficient soybean genotypes and D55,

E108, E150, E277-1 and E283 were chosen to represent the P-inefficient soybean genotypes.

Hydroponic Study

Effect of P Supply on Dry Matter Accumulation

The biomass of soybean genotypes reduced by 6.1~35.1% in low P treatment compared to in high P treatment (Figure 3). Low P level reduced plant biomass by 19.8% averaged the 10 soybean genotypes, but affected the two groups of phenotypes differently. The average shoot biomass, root biomass of P-efficient soybean genotypes was 49.4 and 54.1% higher than that of the P-inefficient genotypes in low P treatment, respectively (Figures 3C,D). Significant difference for shoot and/ or root was also observed among the genotypes in P-efficient and P-inefficient group. The root and shoot biomass of E141, E295, and E311 significantly higher than E14 and E64 in low P condition (Figure 3C). The root biomass of E150 is the highest in P-inefficient group in low P level (Figure 3D).

P Accumulation and Distribution

Low P level reduced P accumulation of soybean genotypes by 32.3~55.8% compared to in high P condition. Averaged P accumulation of 10 soybean genotypes reduced by 46.5% in low P level compare to in high P level (Figure 4). The average shoot P accumulation and root P accumulation of 5 P-efficient soybean genotypes was 17.3 and 22.2% higher than that of P-inefficient genotypes in low P condition, respectively (Figures 4C,D). P accumulation of E311 were 7.5 and 16.9 mg plant⁻¹ in low and high P level respectively, and were 1.4, 1.7 and 1.6, 1.9-folds greater than that of the E64 and E14 grown in media supplied with low and high P concentration, respectively (Figures 4A,C). Shoot P accumulation of E150 was highest among D55, E108, E277-1 and E283, but root P accumulation of the 5 soybean genotypes have no obviously difference (Figures 4B,D).

As shown in Table 1, the ratio of root P accumulation: shoot P accumulation of P-efficient soybean genotypes in low P condition was much high than in high P condition. However, the ratio of the P-inefficient genotypes in low and high P conditions showed narrow difference. For instance, the ratios of E141 and E311 in low P condition were 30.8%, 32.9% and 30.0%, 50.8% higher than in high P level, respectively. But the ratio of E150 in low P condition was 30.0%, and higher than in high P level 9.7%. Low P condition increased the root: shoot ratios of soybean genotypes, but P-efficient soybean genotypes preferentially assimilate P distributed in root.

Root Morphology

Low P reduced root length, root surface area and root volume of soybean genotypes compared with in high P level (Table 2). However, P-efficient soybean genotypes had greater root length, surface area and root volume than P-inefficient genotypes in low and high P condition (Table 2). E141 possessed the highest root length, surface area and root volume in high P level, but those root morphology index decreased by 47.8, 51.2, and 56.4% in the low P level, respectively. On the contrary, root length, root surface area and root volume of E311 in low P level just decreased by 17.4, 16.4, and 2.3% compared with in high P level, respectively

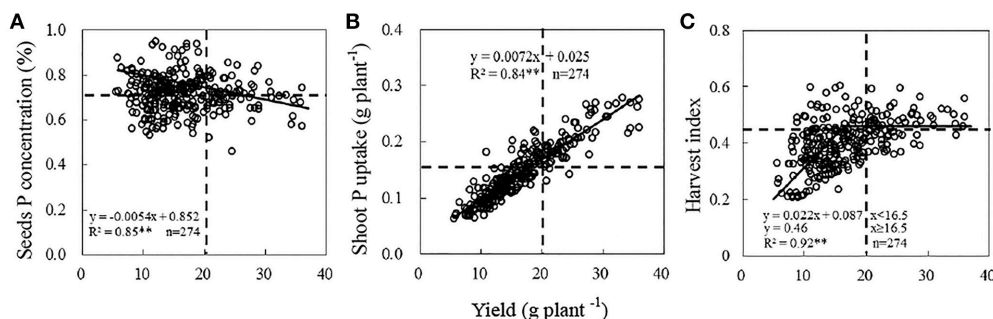


FIGURE 2 | Relationship between soybean yield and seeds P concentration (A), shoot P uptake (B), HI (C). Data are mean of three replicates. The horizontal dotted line in figures (A–C) means soybean seeds P content, shoot P uptake and HI among the genotypes, respectively. The vertical dotted lines in figures (A–C) were representing the soybean yield of *Nandou 12* in Sichuan province (20 g plant^{-1}).

(Table 2). Low P level reduced specific root length of P-inefficient soybean genotypes compared with in high P level, except for E283. But, E64, E141, and E311 reached a higher specific root length (9837, 9381, and $10063 \text{ m g}^{-1} \text{ DW}$, respectively) in low P level than in high P level (Table 2).

Activities of APase

APase activity of root significantly increased in low P level compared to in high P condition (Figure 5). The root APase activities of P-efficient soybean genotypes ranged from 27.8 to $54.21 \text{ pNP ug (g FW)}^{-1} \text{ min}^{-1}$, and the average was $42.1 \text{ pNP ug (g FW)}^{-1} \text{ min}^{-1}$ in low P level (Figure 5A). The corresponding data of P-inefficient soybean genotypes ranging from 20.29 to $33.09 \text{ pNP ug (g FW)}^{-1} \text{ min}^{-1}$ was averaged by $24.1 \text{ pNP ug (g FW)}^{-1} \text{ min}^{-1}$ in low P level (Figure 5B). The maximum APase activities of E311 and E141 reached 52.15 and $54.21 \text{ pNP ug (g FW)}^{-1} \text{ min}^{-1}$ in low P level, respectively, and were 1.2~2.3-fold higher than others (Figure 5A). D55 and E277-1 showed higher APase activities in low P condition than the E108, E150, and E283 in high P level (Figure 5B).

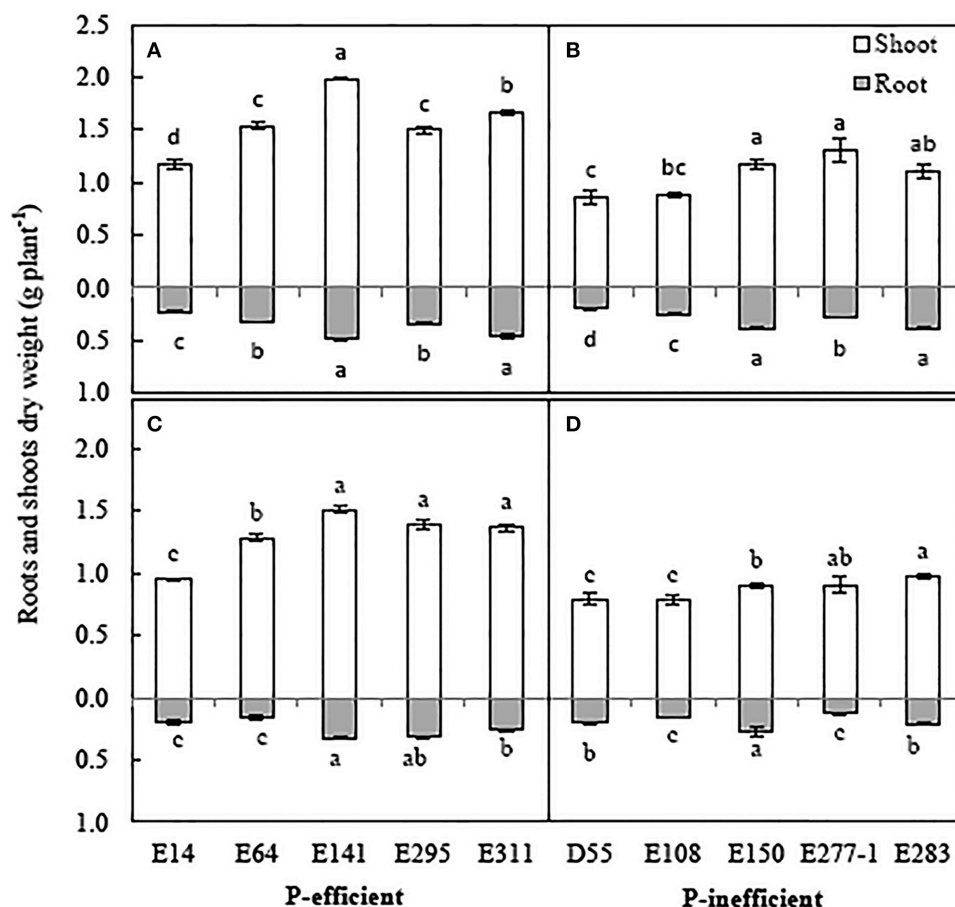
Organic Acid Exudation

Root exudation of malate was greatly motivated in both P-efficient and P-inefficient soybean genotypes by P deficiency (Table 3). The maximum malate exudation rate of 114.6 and $134.8 \text{ mg plant}^{-1} \text{ h}^{-1}$ were found in E311 and E141 in low P level, respectively. Citrate and oxalate exudation of P-efficient and P-inefficient soybean genotypes increased in low P condition. P-efficient soybean genotypes sustained a higher rate of citrate and oxalate efflux than P-inefficient genotypes in low P level. The tartrate efflux was activated by P deficiency and hereditary character. E141 and E311 maintained a high rate of malate, citrate, tartrate, and oxalate excretion, especially in P deficiency condition. Indicating that organic acid exudation by soybean roots was in response to P starvation and genotypes. E150 maintained a high rate of malate, citrate and oxalate excretion than the other four soybean P-efficient genotypes in low P condition. On the contrary, E277-1 and E283 maintained a high rate of tartaric excretion than D55, E108 and E150 in low P level (Table 3).

DISCUSSION

Genetic Variations Existed for Grain Yield and P Accumulation in the Southwest of China

To screen P-efficient soybean genotypes for achieving high yield and improving P use efficiency, 274 soybean genotypes were collected from Southwest of China (Figure 1). Besides low availability of soil P, low P use efficiency partly resulted from the behavior of fertilizer P application by farmers, who often input more P than crops need (Vitousek et al., 2009). The remaining P accumulation in the soil is easily eroded with soil by rainfall (Zhang et al., 2012). To sustainably improve crop production and fertilizer P used efficiency through exploiting the biological P potential in the soil, employing the P-efficient genotypes is an efficacious strategy (Li et al., 2011). In this study, grain yield, shoot P accumulation, seed P concentration and harvest index (HI) of 274 soybean genotypes were analyzed in Purple soil with a suitable soil P concentration (initial Olsen-P concentration was 17.4 mg kg^{-1}). The results showed that seed P concentration was diluted and shoot P accumulation improved following increasing yield (Figure 2). The 274 soybean genotypes were originated from a large span area with low availability of soil P (Figure 1), which differed substantially in grain yield and P accumulation potentials (Figure 2). The yield of 274 soybean genotypes ranged from 5.5 to $36.0 \text{ g plant}^{-1}$. The seed P concentration and shoot P accumulation ranged from 0.045 to 0.93% and $0.065 \text{ mg plant}^{-1}$ to $0.278 \text{ mg plant}^{-1}$, respectively (Figure 2). It suggested that genetic variations existing in the 274 soybean genotypes, and which provide a potential for assessment P-efficient soybean genotypes (Pan et al., 2008). Such substantial genotypic variation in response to different P use efficiency was also detected for a considerable number of soybean genotypes in South and Northeast of China (Zhao et al., 2004; Pan et al., 2008). In the field study, based on the stand enumerated before 5 P-efficient and 5 P-inefficient soybean genotypes were chosen (Figure 2). Obviously, the evidence mentioned above to proving the soybean genotypes with different P use efficiency is not enough, much more work should be provided to support the results as root physiological, chemical



Comparisons for means of two groups of genotypes at high and low P				
Genotypes	High P		Low P	
	shoot	root	shoot	root
Efficient	1.57 a	0.367 a	1.30 a	0.299 a
Inefficient	1.07 b	0.250 b	0.87 b	0.194 b

For means followed by same letter are not significantly different ($p < 0.05$).

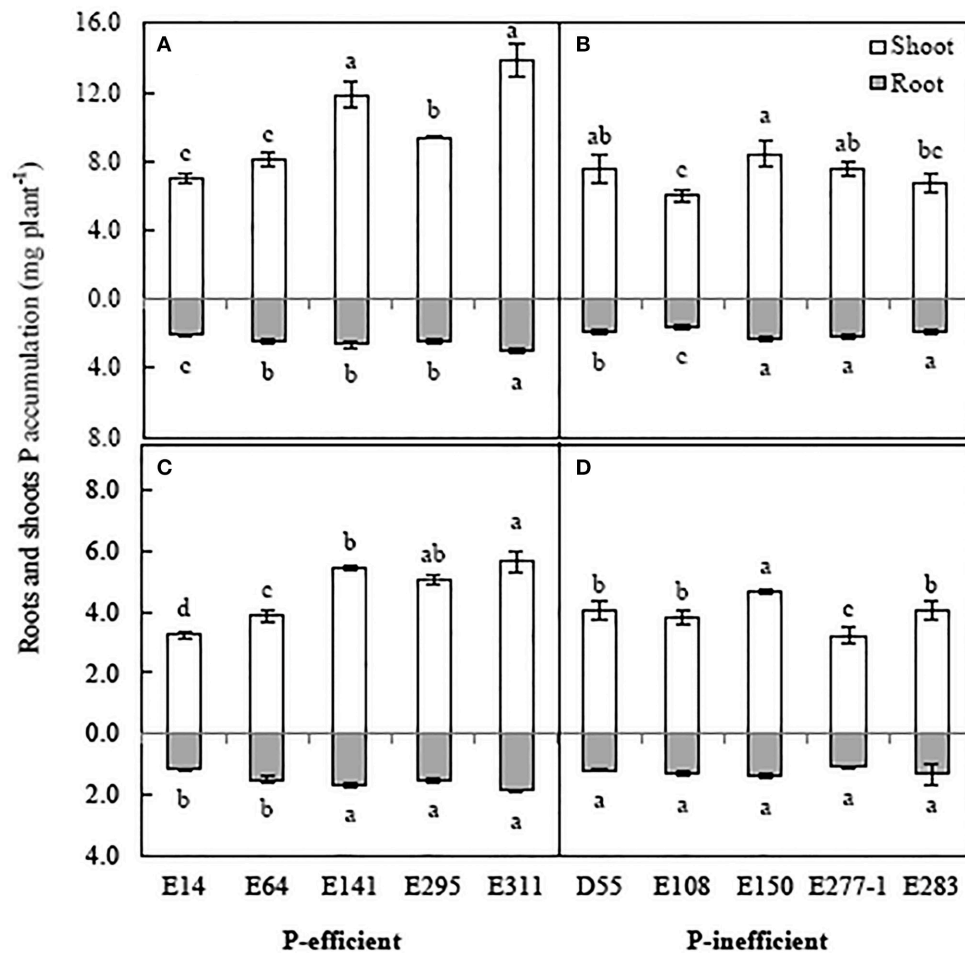
FIGURE 3 | Root and shoot dry weight of soybean genotypes, grown in high and low P conditions in greenhouse. Data are mean of three replicates \pm SE. (A) Biomass of P-efficient soybean genotypes in high P condition. (B) Biomass of P-inefficient soybean genotypes in high P condition. (C) Biomass of P-efficient soybean genotypes in low P condition. (D) Biomass of P-inefficient soybean genotypes in low P condition. Different letters on each column of shoot or root are significantly difference at the 5% level by LSD among soybean genotypes.

characteristics. Hydroponic study is a fine pathway to check plant root characteristics as it is pellucid and efficient by short plant growth period.

Biomass and P Accumulation of Soybean Genotypes

Crop P efficiency was defined as the ability to produce biomass or yield under certain available P supply conditions (Wissuwa et al., 2009), and further explained the capability of P uptake from the media and tolerance in P insufficient conditions (Wissuwa and Ae, 2001; Wissuwa et al., 2005; Wang et al., 2010).

In the hydroponic experiment, P-efficient genotypes exhibited better adaptability and tolerance than the P-inefficient genotypes grown in low P solution (Figure 3C). Especially E141 and E311, showed significantly greater biomass than that of P-inefficient genotypes in low P level (Figure 3). These results corresponded well with previous studies which showed that the P-efficient plant genotypes demonstrated greater biomass compared to the P-inefficient when grown in low P condition. Some scholars had reported that P-efficient plants of soybean (Zhao et al., 2004), rice (*Oryza sativa* L.) (Wissuwa and Ae, 2001; Mori et al., 2016), maize (*Zea mays* L.) (Zhang, 2012), and



Comparisons for means of two groups of genotypes at high and low P

Genotypes	High P		Low P	
	Shoot	root	shoot	root
Efficient	9.94 a	2.65 a	4.94 a	1.38 a
Inefficient	7.47 b	1.77 b	4.09 b	1.14 b

For means followed by same letter are not significantly different ($p < 0.05$).

FIGURE 4 | Root and shoot P accumulation of soybean genotypes, grown in high and low P conditions in greenhouse. Data are mean of three replicates \pm SE. **(A)** P accumulation of P-efficient soybean genotypes in high P condition. **(B)** P accumulation of P-inefficient soybean genotypes in high P condition. **(C)** P accumulation of P-efficient soybean genotypes in low P condition. **(D)** P accumulation of P-inefficient soybean genotypes in low P condition. Different letters on each column of shoot or root are significantly difference at the 5% level by LSD among soybean genotypes.

Brassica napus (Zhang et al., 2009) showed significantly biomass ascendancy compared with P-inefficient plants in P shortage conditions.

Early studies suggested that P-efficient plants preferential distribution P to roots for improving the tolerance to P deficiency and stimulate root growth and P uptake (Wissuwa et al., 2005). In present study, much more P distribution to roots of P-efficient soybean genotypes in low P condition may be an adaptation

involved in increasing the tolerance to P deficiency (Table 1). Tolerant plants have been more efficient in P uptake per root size and the additional P then drove further root growth, assuming that low P availability affected root biomass accumulation directly (Wissuwa et al., 2005). In this study, P accumulation of E141 and E311 was higher than the other genotypes in low P solution (Figure 4C). Some scholars reported that P-efficient soybean genotypes are able to obtain sufficient P from acid red soil and

alkaline soil under P lack conditions (Zhao et al., 2004; Pan et al., 2008), and other P-efficient plant species also showed positively P accumulation grown in P lack conditions, like maize (Liu et al., 2004), wheat (*Triticum aestivum* L.) (Fageria and Baligar, 1999) and rice (Mori et al., 2016). It indicated that the P-efficient soybean genotypes (especially E141 and E311) demonstrate great capability on P absorption and accumulation potentials.

Root Physiological Adaptation Concerned in Stimulate P Assimilation

Root physiological adaptations play important roles in enhancing soil P bioavailability and crop P use efficiency (Shen et al., 2011). These adaptation mechanisms mainly include altering root morphology to enhance P absorption (Wissuwa, 2003),

facilitating organic acids exudation into the rhizosphere to increase P availability by mobilizing sparingly soluble mineral P and organic P sources (Johnson et al., 1996), and promoting phosphatase exudation to mineralize Po (Li et al., 2012). Root morphological traits closely linked with P acquisition ability of plants (Pang et al., 2010), and fine root morphology is propitious to maximize P assimilation. P-efficient soybeans showed high parameters of root length, root surface area and root volume than the P-inefficient in present study (Table 2). The specific root length of E311 increased by 48.6% compared with that in high P level (Table 2). It indicated that penurious P availability has negative effects on root morphology, however the less reduction of those root morphology parameters of P-efficient genotypes seems to be the results of excellent tolerance and adaptation. It reported that P-efficient plant genotypes altered root morphology to adapt low P conditions, and then assimilated more P than P-inefficient genotypes (Fita et al., 2011). There was great possibility for plants in improving soil P excavation and utilization by increasing root length, root surface area, root volume and specific root length (Vance et al., 2003), which result in a large total amount of P and dry matter accumulation (Zhang et al., 2013).

Besides root morphology, exudation of APase to improve P bioavailability is another important root physiological adaptation. As we know, exudation of APase by roots increased under P deficiency condition, and plant roots with high APase activities have great potential to utilize soil Po (Hayes et al., 2000; Lung and Lim, 2006). In this study, the results corresponded well with earlier researches that P shortage conditions stimulate the APase activity of soybean genotypes, especially the P-efficient soybean genotypes (Figure 5). Po comprises 30–80% of the total P in most agricultural soils (Dalai, 1977), it can be

TABLE 1 | P distribution between root and shoot of soybean genotypes grown in low and high P conditions in greenhouse.

Efficient genotypes	Root P accumulation: shoot P accumulation ratio (%)				Inefficient genotypes
	High P	Low P	High P	Low P	
E14	29.8	35.4	25.6	29.6	D55
E64	30.5	39.0	26.8	34.4	E108
E141	22.3	30.8	27.4	30.0	E150
E295	26.6	29.9	28.1	33.2	E277-1
E311	21.8	32.9	27.5	32.5	E283

Data are mean of three replicates. Ten genotypes were separated into two groups (P-efficient group included E14, E64, E141, E295, E311, P-inefficient group included D55, E108, E150, E277-1, E283). The ratio of root P accumulation: shoot P accumulation was counted by root P accumulation divided by shoot P accumulation.

TABLE 2 | Root morphology of soybean genotypes grown in low and high P conditions in greenhouse.

Genotypes	Root length (cm plant ⁻¹)		Root surface area (cm ² plant ⁻¹)		Root volume (cm ³ plant ⁻¹)		Specific root length (cm g ⁻¹ plant ⁻¹)	
	High	Low	High	Low	High	Low	High	Low
P-EFFICIENT								
E14	1926 c*	546 d	279 c*	112 c	3.26 c*	1.31 c	8579 a*	6794 c
E64	3020 b*	1523 c	379 b*	187 bc	3.81 bc*	1.81 c	9350 a	9837 b
E141	3985 a*	2078 b	514 a*	251 b	5.44 a*	2.37 b	8147 a	9381 b*
E295	2313 c*	1354 c	343 c*	246 b	4.01 b*	2.32 b	6853 b*	4248 d
E311	3117 b*	2575 a	469 ab*	392 a	4.21 b	4.12 a	6772 b	10063 a*
P-INEFFICIENT								
D55	1196 d*	568 c	141 c*	75 c	1.36 c*	0.79 c	6057 bc*	2805 c
E108	1392 c*	925 b	196 b*	133 b	2.17 b*	1.49 b	5618 c	5041 b
E150	2451 a*	1584 a	316 a*	219 a	3.48 a*	2.20 a	6400 b*	5801 ab
E277-1	1342 c*	1147 b	221 b*	149 b	1.56 c	1.51 b	7725 a*	5058 b
E283	1660 b*	1386 a	219 b*	165 b	2.29 b*	1.55 b	5325 c	6556 a*
AVERAGE								
Efficient	2872 A*	1615 A	397 A*	238 A	4.15 A*	2.39 A	7940 A	8065 A
Inefficient	1608 B*	1122 B	219 B*	148 B	2.17 B*	1.51 B	6225 B*	5052 B

Data are mean of three replicates. Values followed by the same small letters in each column are not significantly different among different soybean genotypes in one group at the 5% level by LSD. Capital letters in each column are significantly different ($p < 0.05$) between the average of efficient and inefficient group.

*Indicated significantly different ($p < 0.05$) between the two P level.

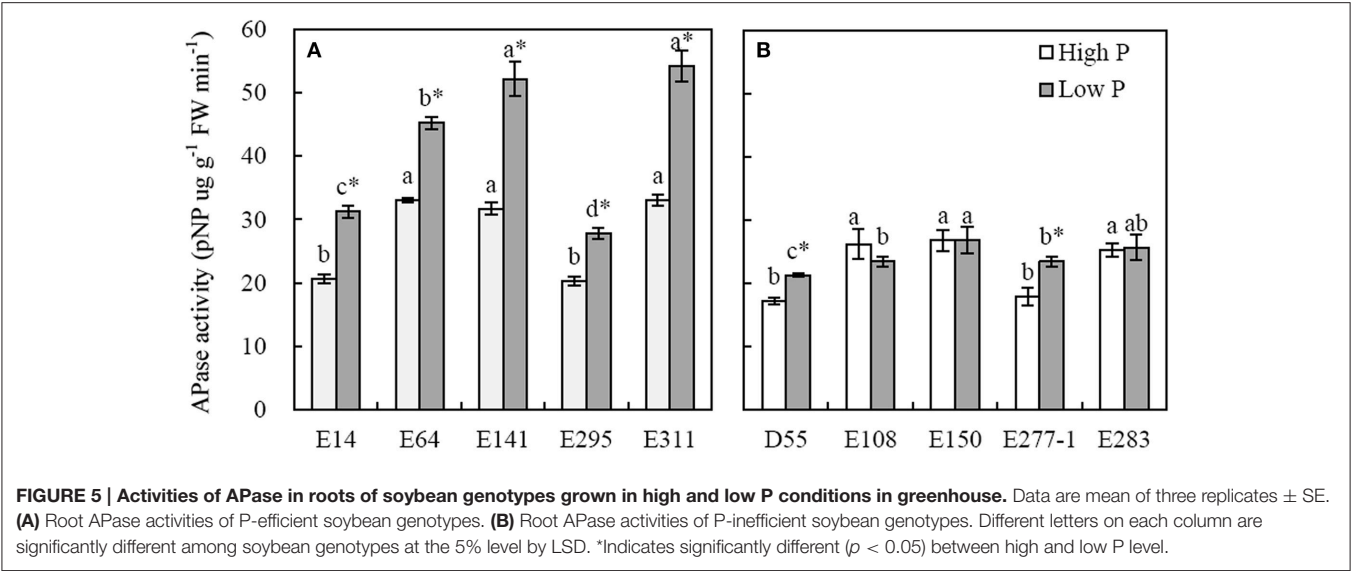


TABLE 3 | Organic acids exudation rate of soybean genotypes grown in low and high P conditions in greenhouse ($\text{mg plant}^{-1} \text{ h}^{-1}$).

Genotypes	Malic acid		Citric acid		Tartaric acid		Oxalic acid	
	High	Low	High	Low	High	Low	High	Low
P-EFFICIENT								
E14	18.8 c	26.3 e*	2.6 c	6.7 bc*	2.3 c	9.4 cd*	0.3 c	2.6 b*
E64	20.8 c	81.2 c*	3.4 b	13.0 a*	11.2 b	12.2 c	0.5 bc	1.5 c*
E141	33.0 b	134.8 a*	4.2 a	13.8 a*	19.5 a	36.4 a*	0.8 b	7.1 a*
E295	24.8 c	37.5 d*	2.6 c	4.8 c	4.3 c	7.9 d*	0.8 b	1.9 c*
E311	77.3 a	114.6 b*	3.7 b	9.4 b*	3.2 c	23.4 b*	2.4 a	3.3 b*
P-INEFFICIENT								
D55	7.3 b	7.2 b	2.7 b	2.0 c	3.4 b	3.4 b	1.5 a*	0.7 c
E108	3.8 c	10.4 b*	2.4 b	2.8 c	3.9 b	4.7 b	0.5 b	0.9 c
E150	18.6 a	52.0 a*	2.8 b	13.2 a*	7.2 a	5.5 b	0.5 b	2.3 a*
E277-1	12.5 b	51.6 a*	1.9 c	8.3 b*	6.3 a	9.5 a*	0.5 b	1.4 b*
E283	24.4 a	54.6 a*	3.5 a	7.6 b*	3.9 b	7.0 a*	1.2 a	1.4 b
AVERAGE								
Efficient	34.9 A	78.9 A*	3.3 A	9.5 A*	8.1 A	17.9 A*	1.0 A	3.3 A*
Inefficient	13.3 B	35.2 B*	2.6 B	6.8 B*	4.9 B	6.0 B	0.8 B	1.3 B*

Data are mean of three replicates. Values followed by the same small letters in each column are not significantly different among different soybean genotypes in one group at the 5% level by LSD. Capital letters in each column are significantly different ($p < 0.05$) between the average of efficient and inefficient group. *Indicated significantly different ($p < 0.05$) between the two P level.

released through mineralization processes mediated by enzymes activity. The P-efficient plant species generally had high enzymes activity in root extracts when grown in P lack or high Po media. For instance, maize, lupin and chickpea showed high APase activities when grown in P deficiency conditions (Wasaki et al., 2003; Li et al., 2004), *Gulf*, *Marshall Ryegrass* and *P. hydropiper* had greater APase activities when grown in phytate-sufficient media (Sharma and Sahi, 2011; Huang et al., 2012). In our results, the APase activities of E311 and E141 were significantly higher (1.2~2.3-folds) than others in low P level (Figure 5C), suggesting that they demonstrated great capabilities of activation and utilization soil P. Besides, transgenic expression

of APase gene in P-efficient genotypes was another path for great dry matter and P accumulation. It reported that over-expressed *AtPAP15* in soybean and *GmPAP4* in *Arabidopsis* significantly increased dry matter and P accumulation compared to the wild type when grown in media with Po (Wang et al., 2009; Kong et al., 2014). Above all, APase activity was a symbol of efficient mineralization and utilization soil P by plants. Root exudation of organic acids into the rhizosphere had been proposed to improve soil P availability and plant P accumulation (Dinkelaker et al., 1989; Johnson et al., 1996). Exudation of organic acids is also an important root physiological adaptation

to P deficiency. A 2-fold increase in exudation of citrate was observed under P starvation in alfalfa (Lipton et al., 1987). Dong et al. (2004) also reported that oxalate and malate exudation of soybean plants was found to markedly increase in response to P deficiency. In present study, low P condition stimulates exudation of malate, citrate, tartrate and oxalate acid, especially in the P-efficient soybean genotypes (Table 3). The average exudation rate of malate, citrate, tartrate and oxalate acid in the P-efficient soybean genotypes increased by 2.24, 1.40, 2.98, and 2.54 folds compared to the P-inefficient genotypes under P starvation, respectively. Particularly, E141 and E311 exhibited high organic acid exudation potentials (Table 3). These results were in agreement with Dong et al. (2004), who reported that a considerable amount of organic acid exudation in P-efficient soybean genotypes contributed to P accumulation under P starvation condition. Adequate evidences demonstrated that root exudation of organic acids into the soil rhizosphere contributed to higher P acquisition. Exudation of malate in P-efficient faba bean enhanced P acquisition in the alkaline soil (Rose et al., 2010). Barley (*Hordeum vulgare* L.) genotypes expressing the *TaALMT1* gene from wheat improved P uptake per unit root length from an acid soil (Delhaize et al., 2009).

CONCLUSION

Field study provided evidence that P use efficiency difference exist in 274 soybean genotypes. The hydroponic study suggested that the P-efficient genotypes are more tolerant to P deficiency, and exhibited fine root morphology and physiological adaptations. In conclusion, our results demonstrated that genotypic variation on P use efficiency existed in yield, P accumulation potentials and root physiological characteristics in

the Southwest of China by field and hydroponic experiments. We suggested that P-efficient soybean genotypes, like E311 and E141, with fine root morphology, high level of root APase activities and exudation of organic acids rate could be potential materials in breeding and improving the production and P use efficiency in intensive agricultural region with low soil P availability. Interestingly, P-inefficient genotype E150 gained low yield in the field but showed fine root morphology and physiological adaptations in the hydroponic experiment, further studies are needed to probe the reason of this phenomenon.

AUTHOR CONTRIBUTIONS

TZ, WL, and WY carried out the design of this research work and writing this paper. TZ, YD, and SA carried out the plant cultivation, chemical analysis and statistical analysis of this work. TL and MR participated in experiment management.

ACKNOWLEDGMENTS

This study was carried out with support from the National Natural Science Foundation of China (31671626) and National Key Research and development Program of China (2016YFD0300209). The authors also wish to thank Jiang Liu for provide the analysis method of organic acids. Besides, Li Wang, Jianyang Gan participated in the chemical analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01776/full#supplementary-material>

REFERENCES

- Cong, W. F., Hoffland, E., Li, L., Six, J., Bao, X. G., Zhng, F. S., et al. (2015). Intercropping enhances soil carbon and nitrogen. *Global Change Biol.* 21, 1715–1726. doi: 10.1111/gcb.12738
- Cordell, D., Drangert, J. O., and White, S. (2009). The story of phosphorus: global food security and food for thought. *Global Environ. Change* 19, 292–305. doi: 10.1016/j.gloenvcha.2008.10.009
- Dalai, R. C. (1977). Soil organic phosphorus. *Adv. Agron.* 29, 83–117. doi: 10.1016/S0065-2113(08)60216-3
- Delhaize, E., Taylor, P., Hocking, P. J., Simpson, R. J., Ryan, P. R., and Richardson, A. E. (2009). Transgenic barley (*Hordeum vulgare* L.) expressing the wheat aluminium resistance gene (*TaALMT1*) shows enhanced phosphorus nutrition and grain production when grown on an acid soil. *Plant Biotechnol. J.* 7, 391–400. doi: 10.1111/j.1467-7652.2009.00403.x
- Dinkelaker, B., Römhild, V., and Marschner, H. (1989). Citric acid excretion and precipitation of calcium citrate in the rhizosphere of white lupin (*Lupinus albus* L.). *Plant Cell Environ.* 12, 285–292. doi: 10.1111/j.1365-3040.1989.tb01942.x
- Dong, D. D., Peng, X. X., and Yan, X. L. (2004). Organic acid exudation induced by phosphorus deficiency and/or aluminium toxicity in two contrasting soybean genotypes. *Physiol. Plant.* 122, 190–199. doi: 10.1111/j.1399-3054.2004.00373.x
- Fageria, N. K., and Baligar, V. C. (1999). Phosphorus-use efficiency in wheat genotypes. *J. Plant Nutr.* 22, 331–340. doi: 10.1080/01904169909365630
- Fageria, N. K., Baligar, V. C., and Li, Y. C. (2008). The role of nutrient efficient plants in improving crop yields in the twenty first century. *J. Plant Nutr.* 31, 1121–1157. doi: 10.1080/01904160802116068
- FAO (2011). *FAOSTAT Database-Agriculture Production*. Rome: Food and Agriculture Organization of the United Nations.
- Fita, A., Nuez, F., and Picó, B. (2011). Diversity in root architecture and response to P deficiency in seedlings of *Cucumis melo* L. *Euphytica* 181, 323–339. doi: 10.1007/s10681-011-0432-z
- Gao, X., Wu, M., Xu, R. N., Wang, X. R., Pan, R. Q., Kim, H. J., et al. (2014). Root interactions in a maize/soybean intercropping system control soybean soil-borne disease, red crown rot. *PLoS ONE* 9:e95031. doi: 10.1371/journal.pone.0095031
- Gregory, P. J., and Atwell, B. J. (1991). The fate of carbon in pulse-labelled crops of barley and wheat. *Plant Soil* 136, 205–213. doi: 10.1007/BF02150051
- Hayes, J. E., Richardson, A. E., and Simpson, R. J. (2000). Components of organic phosphorus in soil extracts that are hydrolysed by phytase and acid phosphatase. *Biolo. Fertil. Soils* 32, 279–286. doi: 10.1007/s003740000249
- Huang, X., Li, T. X., Zhang, X. Z., Zheng, Z. C., and Yu, H. Y. (2012). Growth, P accumulation, and physiological characteristics of two ecotypes of *Polygonum hydropiper* affected by excess P supply. *J. Plant Nutr. Soil Sci.* 175, 290–302. doi: 10.1002/jpln.201100067
- Hymowitz, T. (1970). On the domestication of the soybean. *Econ. Bot.* 24, 408–421. doi: 10.1007/BF02860745
- Johnson, J. F., Vance, C. P., and Allan, D. L. (1996). Phosphorus deficiency in *Lupinus albus*. Altered lateral root development and enhanced expression of phosphoenolpyruvate carboxylase. *Plant Physiol.* 112, 31–41. doi: 10.1104/pp.112.1.31
- Kong, Y. B., Li, X. H., Ma, J., Li, W. L., Yan, G. J., and Zhang, C. Y. (2014). GmPAP4, a novel purple acid phosphatase gene isolated from soybean (*Glycine max*),

- enhanced extracellular phytate utilization in *Arabidopsis thaliana*. *Plant Cell Rep.* 33, 655–667. doi: 10.1007/s00299-014-1588-5
- Koyama, H., Kawamura, A., Kihara, T., Hara, T., Takita, E., and Shibata, D. (2000). Overexpression of mitochondrial citrate synthase in *Arabidopsis thaliana* improved growth on a phosphorus-limited soil. *Plant Cell Physiol.* 41, 1030–1037. doi: 10.1093/pcp/pcd029
- Li, C. C., Gui, S. H., Yang, T., Walk, T., Wang, X. R., and Liao, H. (2012). Identification of soybean purple acid phosphatase genes and their response to phosphorus availability and symbiosis. *Ann. Bot. Lond.* 109, 275–285. doi: 10.1093/aob/mcr246
- Li, H., Huang, G., Meng, Q., Ma, L., Yuan, L., Wang, F., et al. (2011). Integrated soil and plant phosphorus management for crop and environment in China. *Rev. Plant Soil* 349, 157–167. doi: 10.1007/s11104-011-0909-5
- Li, S. M., Li, L., Zhang, F. S., and Tang, C. (2004). Acid phosphatase role in chickpea/maize intercropping. *Ann. Bot. Lond.* 94, 297–303. doi: 10.1093/aob/mch140
- Libert, B., and Franceschi, V. R. (1987). Oxalate in crop plants. *J. Agric. Food Chem.* 35, 926–938. doi: 10.1021/jf00078a019
- Lin, C. W., Tu, S. H., Huang, J. J., and Chen, Y. B. (2009). The effect of plant hedgerows on the spatial distribution of soil erosion and soil fertility on sloping farmland in the purple-soil area of China. *Soil Till. Res.* 105, 307–312. doi: 10.1016/j.still.2009.01.001
- Lipton, D. S., Blanchar, R. W., Blevins, D. G., and Blevins, D. G. (1987). Citrate, malate and succinate concentration in exudates from P-sufficient and P-stressed *Medeicago sativa* L. seedlings. *Plant Physiol.* 85, 315–317. doi: 10.1104/pp.85.2.315
- Liu, W. G., Zou, J. L., Zhang, J., Yang, F., Yan, W., and Yang, W. Y. (2015). Evaluation of soybean (*Glycine max*) stem vining in maize-soybean relay strip intercropping system. *Plant Prod. Sci.* 18, 69–75. doi: 10.1626/pp.18.69
- Liu, Y., Mi, G. H., Chen, F. J., Zhang, J. H., and Zhang, F. S. (2004). Rhizosphere effect and root growth of two maize (*Zea mays* L.) genotypes with contrasting P efficiency at low P availability. *Plant Sci.* 167, 217–223. doi: 10.1016/j.plantsci.2004.02.026
- Lung, S. C., and Lim, B. L. (2006). Assimilation of phytate-phosphorus by the extracellular phytase activity of tobacco (*Nicotiana tabacum*) is affected by the availability of soluble phytate. *Plant Soil* 279, 187–199. doi: 10.1007/s11104-005-1009-1
- Mori, A., Fukuda, T., Vejchasarn, P., Nestler, J., Pariasca-Juan, T., and Wissuwa, M. (2016). The role of root size versus root efficiency in phosphorus acquisition of rice. *J. Exp. Bot.* 67, 991–993. doi: 10.1093/jxb/erv557
- Otani, T., and Ae, N. (1996). Phosphorus (P) uptake mechanisms of crops grown in soils with low P status I. Screening of crops for efficient P uptake. *Soil Sci. Plant Nutr.* 42, 155–163. doi: 10.1080/00380768.1996.10414699
- Page, A. L. (1982). *Methods of Soil Analysis (Part 2)*, 2nd Edn. Madison, WI: American Society of Agronomy.
- Pan, X. W., Li, W. B., Zhang, Q. Y., Li, Y. H., and Liu, M. S. (2008). Assessment on phosphorus efficiency characteristics of soybean genotypes on phosphorus-deficient soils. *Agr. Sci. China* 7, 958–969. doi: 10.1016/S1671-2927(08)60135-2
- Pang, J., Ryan, M. H., Tibbett, M., Cawthray, G. R., Siddique, K. H., Bolland, M. D., et al. (2010). Variation in morphological and physiological parameters in herbaceous perennial legumes in response to phosphorus supply. *Plant Soil* 331, 241–255. doi: 10.1007/s11104-009-0249-x
- Rose, T. J., Damon, P., and Rengel, Z. (2010). Phosphorus-efficient faba bean (*Vicia faba* L.) genotypes enhance subsequent wheat crop growth in an acid and an alkaline soil. *Crop Past. Sci.* 61, 1009–1016. doi: 10.1071/CP10205
- Salviaggiotti, F., Cassman, K. G., Specht, J. E., Walters, D. T., Weiss, A., and Dobermann, A. (2008). Nitrogen uptake, fixation and response to fertilizer N in soybeans: a review. *Field Crops Res.* 108, 1–13. doi: 10.1016/j.fcr.2008.03.001
- Sharma, N. C., and Sahi, S. V. (2011). Enhanced organic phosphorus assimilation promoting biomass and shoot Phyper accumulations in *Lolium multiflorum* grown under sterile conditions. *Environ. Sci. Technol.* 45, 10531–10537. doi: 10.1021/es200942v
- Sharma, N. C., and Sahi, S. V. (2005). Characterization of phosphate accumulation in *Lolium multiflorum* for remediation of phosphorus-enriched soils. *Environ. Sci. Technol.* 39, 5475–5480. doi: 10.1021/es050198t
- Shen, J., Yuan, L., Zhang, J. L., Li, H., Bai, Z., Chen, X., et al. (2011). Phosphorus dynamics: from soil to plant. *Plant Physiol.* 156, 997–1005. doi: 10.1104/pp.111.175232
- Vance, C. P., Uhde, S. C., and Allan, D. L. (2003). Phosphorus acquisition and use: critical adaptations by plants for securing a nonrenewable resource. *New Phytol.* 157, 423–447. doi: 10.1046/j.1469-8137.2003.00695.x
- Vincent, J. M. (1970). *A Manual for the Practical Study of the Root-Nodule Bacteria*. Oxford: Blackwell Scientific.
- Vitousek, P. M., Naylor, R., Crews, T., David, M. B., Drinkwater, L. E., Holland, E., et al. (2009). Nutrient imbalances in agricultural development. *Science* 324, 1519–1520. doi: 10.1126/science.1170261
- Wang, X. R., Shen, J. B., and Liao, H. (2010). Acquisition or utilization, which is more critical for enhancing phosphorus efficiency in modern crop? *Plant Sci.* 179, 302–306. doi: 10.1016/j.plantsci.2010.06.007
- Wang, X., Wang, Y., Tian, J., Lim, B., Yan, X., and Liao, H. (2009). Overexpressing *AtPAP15* enhances phosphorus efficiency in soybean. *Plant Physiol.* 151, 233–240. doi: 10.1104/pp.109.138891
- Wasaki, J., Maruyama, H., Tanaka, M., Yamamura, T., and Dateki, H., Shinano, T. et al. (2009). Overexpression of the LASAP2 gene for secretory acid phosphatase in white lupin improves the phosphorus uptake and growth of tobacco plants. *Soil Sci. Plant Nutr.* 55, 107–113. doi: 10.1111/j.1747-0765.2008.00329.x
- Wasaki, J., Yamamura, T., Shinano, T., and Osaki, M. (2003). Secreted acid phosphatase is expressed in cluster roots of lupin in response to phosphorus deficiency. *Plant Soil* 248, 129–136. doi: 10.1023/A:1022332320384
- Wissuwa, M. (2003). How do plants achieve tolerance to phosphorus deficiency? Small causes with big effects. *Plant Physiol.* 133, 1947–1958. doi: 10.1104/pp.103.029306
- Wissuwa, M., and Ae, N. (2001). Genotypic variation for tolerance to phosphorus deficiency in rice and the potential for its exploitation in rice improvement. *Plant Breed.* 120, 43–48. doi: 10.1046/j.1439-0523.2001.00561.x
- Wissuwa, M., Gamat, G., and Ismail, A. M. (2005). Is root growth under phosphorus deficiency affected by source or sink limitation? *J. Exp. Bot.* 56, 1943–1950. doi: 10.1093/jxb/eri189
- Wissuwa, M., Mazzola, M., and Picard, C. (2009). Novel approaches in plant breeding for rhizosphere-related traits. *Plant Soil* 321, 409–430. doi: 10.1007/s11104-008-9693-2
- Xia, H. Y., Wang, Z. G., Zhao, J. H., Sun, J. H., Bao, X. G., and Christie, P. (2013). Contribution of interspecific interactions and phosphorus application to sustainable and productive intercropping systems. *Field Crops Res.* 154, 53–64. doi: 10.1016/j.fcr.2013.07.011
- Yang, F., Huang, S., Gao, R. C., Liu, W. G., Yong, T. W., Wang, X. C., et al. (2014). Growth of soybean seedlings in relay strip intercropping systems in relation to light quantity and red:far-red ratio. *Field Crops Res.* 155, 245–253. doi: 10.1016/j.fcr.2013.08.011
- Yu, L., Peng, X. X., Yang, C., Liu, Y. H., and Fan, Y. P. (2002). Determination of oxalic acid in plant tissue and root exudates by reversed phase high performance liquid chromatography. *Chin. J. Anal. Chem.* 30, 1119–1122. doi: 10.3321/j.issn:0253-3820.2002.09.027
- Zhang, F. S., Cui, Z. L., Chen, X. P., Ju, X. T., Shen, J. B., Chen, Q., et al. (2012). Integrated nutrient management for food security and environmental quality in China. *Adv. Agron.* 116, 1–40. doi: 10.1016/B978-0-12-394277-7.00001-4
- Zhang, H. W., Huang, Y., Ye, X. S., Shi, L., and Xu, F. S. (2009). Genotypic differences in phosphorus acquisition and the rhizosphere properties of *Brassica napus* in response to low phosphorus stress. *Plant Soil* 320, 91–102. doi: 10.1007/s11104-008-9873-0
- Zhang, W. F., Ma, W. Q., Ji, Y. X., Fan, M. S., Oenema, O., and Zhang, F. S. (2008). Efficiency, economics, and environmental implications of phosphorus resource use and the fertilizer industry in China. *Nutr. Cycl. Agroecosyst.* 80, 131–144. doi: 10.1007/s10705-007-9126-2
- Zhang, X. B., He, X. B., Wen, A. B., Walling, D. E., Feng, M. Y., and Zou, X. (2004). Sediment source identification by using ¹³⁷Cs and ²¹⁰Pb radionuclides in a small catchment of the Hilly Sichuan Basin, China. *Chinese Sci. Bull.* 49, 1953–1957. doi: 10.1007/BF03184288
- Zhang, Y. K. (2012). *Genetic Improvement of Root Growth Contributes to Efficient Phosphorus Acquisition in Maize (Zea mays L.)*. Dissertation for Doctoral Degree. China Agricultural University, Beijing.

- Zhang, Y. K., Chen, F. J., Chen, X. C., Long, L. Z., Gao, K., Yuan, L. X., et al. (2013). Genetic improvement of root growth contributes to efficient phosphorus acquisition in maize (*Zea mays* L.). *J. Integr. Agric.* 12, 1098–1111. doi: 10.1016/S2095-3119(13)60489-X
- Zhao, J., Fu, J. B., Liao, H., He, Y., Nian, H., Hu, Y. M., et al. (2004). Evaluation for root configuration traits in phosphorus efficiency of soybean applying core soybean germplasm. *Chin. Sci. Bull.* 13, 1249–1257. doi: 10.3321/j.issn:0023-074X.2004.13.006

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhou, Du, Ahmed, Liu, Ren, Liu and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Finger on the Pulse: Pumping Iron into Chickpea

Grace Z. H. Tan¹, Sudipta S. Das Bhowmik¹, Thi M. L. Hoang¹, Mohammad R. Karbaschi¹, Alexander A. T. Johnson², Brett Williams¹ and Sagadevan G. Mundree^{1*}

¹ Centre for Tropical Crops and Biocommodities, Queensland University of Technology, Brisbane, QLD, Australia, ² School of Biosciences, University of Melbourne, Melbourne, VIC, Australia

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica - Universidade Nova de
Lisboa, Portugal

Reviewed by:

María Ayelén Pagani,
Consejo Nacional de Investigaciones
Científicas y Técnicas (CONICET),
Argentina
Ambuj Bhushan Jha,
University of Saskatchewan, Canada

*Correspondence:

Sagadevan G. Mundree
sagadevan.mundree@qut.edu.au

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 30 December 2016

Accepted: 25 September 2017

Published: 13 October 2017

Citation:

Tan GZH, Das Bhowmik SS,
Hoang TML, Karbaschi MR,
Johnson AAT, Williams B and
Mundree SG (2017) Finger on the
Pulse: Pumping Iron into Chickpea.
Front. Plant Sci. 8:1755.
doi: 10.3389/fpls.2017.01755

Iron deficiency is a major problem in both developing and developed countries, and much of this can be attributed to insufficient dietary intake. Over the past decades several measures, such as supplementation and food fortification, have helped to alleviate this problem. However, their associated costs limit their accessibility and effectiveness, particularly amongst the financially constrained. A more affordable and sustainable option that can be implemented alongside existing measures is biofortification. To date, much work has been invested into staples like cereals and root crops—this has culminated in the successful generation of high iron-accumulating lines in rice and pearl millet. More recently, pulses have gained attention as targets for biofortification. Being secondary staples rich in protein, they are a nutritional complement to the traditional starchy staples. Despite the relative youth of this interest, considerable advances have already been made concerning the biofortification of pulses. Several studies have been conducted in bean, chickpea, lentil, and pea to assess existing germplasm for high iron-accumulating traits. However, little is known about the molecular workings behind these traits, particularly in a leguminous context, and biofortification via genetic modification (GM) remains to be attempted. This review examines the current state of the iron biofortification in pulses, particularly chickpea. The challenges concerning biofortification in pulses are also discussed. Specifically, the potential application of transgenic technology is explored, with focus on the genes that have been successfully used in biofortification efforts in rice.

Keywords: pulse biofortification, iron, genetic modification, crop improvement, chickpea

INTRODUCTION

The current world population stands at an estimated 7.3 billion (United Nations, 2015) and is projected to increase by 2 billion over the next four decades. Concomitant to this growth is the challenge of providing sustenance amidst dwindling resources. Currently food production is adequate at approximately four billion metric tons per annum, yet in spite of this, about 870 million people still suffer from chronic malnutrition due to factors like unequal distribution, wastage and poor diets (FAO, 2012; IMECHE, 2013).

Malnutrition, as defined by the World Health Organization (WHO), is “the cellular disparity amid the supply of energy, nutrients and the body’s demand for them to ascertain maintenance, growth and specific functions” (Batool et al., 2013). It refers to both the insufficient and excessive intake of nutrients (both macro and micro) and as such covers not only food shortage but

also obesity. Undernourishment can be classified categories: protein-energy malnutrition and micronutrient deficiency. As the names suggest, the former refers to inadequate calorie or protein intake while the latter to the lack of essential micronutrients such as vitamin A, iodine, zinc, and iron (Batool et al., 2013).

While both pose significant risks to health and negatively affect overall productivity and quality of life, micronutrient deficiency, also known “hidden hunger,” is perhaps the more pervasive and lethal due to the lack of visible effects. It is consequently more difficult to identify and tackle, and afflicts both developing and developed nations.

Among the various kinds of micronutrient deficiencies, iron deficiency is the most prevalent, afflicting more than two billion individuals worldwide (WHO, 2008). It has been identified as the greatest contributor to anemia, accounting for 66.2% of cases globally (Alvarez-Uria et al., 2014). The extent of its impact is such that the terms are used interchangeably and the prevalence of anemia is used as a measure for the more specific iron deficiency anemia (IDA) (WHO, 2001). IDA can be attributed to three main factors—increased iron requirement (e.g., growth and pregnancy), poor absorption, and inadequate dietary intake. The recommended values for daily iron intake varies depending on the gender and developmental stage (**Table 1**), and insufficient intake impedes the formation of biologically important compounds, most notably heme, resulting in anemia. Symptoms include fatigue, loss of energy, and dizziness, all of which diminish the work capacity of the individual. Iron deficiency also results in poor pregnancy outcomes and impediment of physical and cognitive development, thereby increasing the risk of morbidity in children (WHO, 2008).

This presents a problem of great economic and social significance, particularly in developing countries where approximately 50% of pregnant women and 40% of preschool children suffer from IDA (WHO, 2008). The consequences manifest not only in the form of lives lost, but also in a rising generation of individuals afflicted with developmental complications. The significance of this issue has been understood by various governments, and through both nutrition and non-nutrition based interventions, considerable progress has been made in reducing IDA, particularly in South Asia, East Asia, Southeast Asia, and Eastern sub-Saharan Africa (Kassebaum, 2016). However, while a global decrease in the prevalence of severe anemia cases was observed between 1990 and 2013, the number of mild to moderate cases has increased (**Table 2**), and a prevalence rate below 10% has yet to be seen in any country (Kassebaum, 2016).

Out of the three major risk factors contributing to IDA, the issue of dietary intake is the most feasible to address on a large scale. Efforts to remedy the problem include changes in policy, education, and food-based strategies. The latter can, in turn, come in various forms such as dietary diversification, food fortification, and supplementation, the definitions and examples of which are illustrated in **Table 3**. More specific details and an overview of the strategies can be found in the published guidelines by the WHO and FAO (2006).

TABLE 1 | Recommended Dietary Allowances (RDAs) for iron (Trumbo et al., 2001).

Age	Male (mg)	Female (mg)	Pregnancy (mg)	Lactation (mg)
Birth to 6 months		0.27		
7–12 months		11		
1–3 years		7		
4–8 years		10		
9–13 years		8		
14–18 years	11	15	27	10
19–50 years	8	18	27	9
51+ years		8		

To summarize, each strategy has its own unique advantages and several studies have proven their effectiveness in alleviating IDA (Baltussen et al., 2004; Gera et al., 2012; Rao et al., 2013). While that efficacy may vary across different temporal and spatial scales, the strategies can be used in concert to greater effect. For instance, dietary diversification can still be used where supplementation or food fortification strategies may not due to lack of suitable infrastructure or distribution networks. However, it in turn, is subject to local environmental conditions and resource availability.

Concerning the aforementioned strategies and their application, the issue of accessibility has been noted to be a major limitation (WHO and FAO, 2006). With the food fortification and supplementation schemes in particular, the recurring costs associated with processing and distribution can be prohibitive and beneficiaries are limited to those who can afford it. Such measures are therefore unfeasible for the low-income demographics that, incidentally, have the greatest need. The challenge then is to develop a sustainable, cost-effective means to deliver the required nutrients to the vulnerable parties.

One such means is biofortification, which can generally be defined as the enhancement of nutritional quality in the edible portions of food crops during plant growth (HarvestPlus, 2015b; WHO, 2016). Given that the process of plant nutrient accumulation is a complex interplay between genetics, environmental and management factors, the precise definition of the term “biofortification” may vary depending on the scope of the means (HarvestPlus, 2015b; WHO, 2016). For the purpose of this review which focuses on the genetic aspect, the term “biofortification” shall be used to refer solely to the generation of self-fortifying plants, to the exclusion of agronomic interventions. Such agronomic interventions include fertilizer application or bacterial inoculation, which can be used in conjunction with biofortified crops. Fertilizer application has been demonstrated to increase iron accumulation, though the degree of which varies between studies (Pahlavan-Rad and Pessarakli, 2009; Cakmak et al., 2010; Zhang et al., 2010; Aciksoz et al., 2011). The use of bacterial inoculation on the other hand, has been met with some success, though the choice of strains used may depend on the environmental conditions (Mishra et al., 2011; Rana et al., 2012; Sharma et al., 2013).

TABLE 2 | Prevalence of anemia between 1990 and 2013 (Kassebaum, 2016).

	Severity of anemia	Prevalent cases of anemia		Difference between 1990 and 2013 (%)
		1990	2013	
Global	Mild	839,101,225	950,135,191	13.23
	Moderate	901,120,023	905,501,751	0.49
	Severe	88,720,928	75,565,628	−14.83
Developing countries	Mild	717,671,655	816,531,244	13.78
	Moderate	803,792,125	809,200,194	0.67
	Severe	82,149,100	69,480,441	−15.42
Developed countries	Mild	121,429,570	133,603,946	10.03
	Moderate	97,327,897.80	96,301,556.30	−1.05
	Severe	6,571,827.50	6,085,186.70	−7.40

TABLE 3 | The main food-based strategies to combat iron deficiency (WHO and FAO, 2006).

	Dietary diversification	Food fortification	Supplementation
Definition	Inclusion and increased intake of foods rich in the target nutrient	Improvement of food nutritional quality via enhancing target nutrient content	Intake of moderately large doses of target nutrient. Can be done via oral (e.g., pills, capsules, or syrups) or intravenous routes
Examples	<ul style="list-style-type: none"> • Leafy vegetables • Pulses • Red meat • Seafood 	<ul style="list-style-type: none"> • Iron-fortified cereals, flour, and bread 	<ul style="list-style-type: none"> • Ferrous salts (e.g., ferrous sulfate, ferrous gluconate, ferrous fumarate) • Iron dextran • Iron sucrose

BIOFORTIFICATION AS A MEANS OF ALLEVIATING GLOBAL IRON DEFICIENCY

Biofortification emerged within the last two decades as an approach to combat micronutrient deficiency. While it cannot be considered a cure-all to micronutrient deficiency, it alleviates the problem by complementing existing strategies like the aforementioned ones of dietary diversification, fortification, and supplementation. With the one-time cost of development thoroughly compensated by the long term benefits, biofortification presents a sustainable means of delivering the needed micronutrients across large spatial and temporal scales (Nestel et al., 2006; Horton et al., 2008; De Moura et al., 2014; HarvestPlus, 2015b). Biofortified crops are typically generated via selection of micronutrient accumulating traits, and there are a few means through which this can be achieved. Amongst these, one that has existed since the advent of agriculture is conventional breeding. Traditionally a long-term process requiring much investment of time and effort, advances in technology and molecular biology has since shortened the process and increased its precision when targeting specific traits. Several quantitative trait loci (QTLs) for iron accumulation have been identified in rice (Norton et al., 2010; Anuradha et al., 2012), wheat (Xu et al., 2012), maize (Jin et al., 2013), pearl millet (Kumar et al., 2016), cowpea (Santos and Boiteux, 2015), and bean (Blair et al., 2009, 2010, 2016). Already, several crops have been developed through conventional breeding under

the HarvestPlus program, the most notable of which are iron biofortified pearl millet, rice and beans. The success of these biofortified crops has been demonstrated in several feeding trials. Consumption of biofortified pearl millet improved iron adsorption and iron stores in women and children (Cercamondi et al., 2013; Kodkany et al., 2013; Finkelstein et al., 2015), while biofortified rice have been found to help maintain the iron stores of non-anemic women (Haas et al., 2005). Increased iron absorption was also observed in biofortified bean meals (Petry et al., 2014, 2016). Collectively, meta-analysis of these trials indicated such biofortified crops to be particularly beneficial to iron deficient individuals (Finkelstein et al., 2017).

Despite its effectiveness, the extent to which biofortification can be done through conventional breeding is limited to the diversity in the gene pool and fertility of the species. In cases where such limitations prevail, genetic modification (GM) provides an alternative pathway. In this method, the genetic material of the host is altered in a manner that does not occur naturally. This may take the form of overexpression of a native gene, such as the OsNAS gene family in rice (Johnson et al., 2011), or expression of a foreign gene from an external source, such as the algal *FEA1* gene in cassava (Ihemere et al., 2012). A major advantage of GM is its specificity—select genes, and thus related traits, can be introduced without linkage drag that is associated with unfavorable agronomic traits. Depending on the gene combinations used, increases in iron content of up to 7.5-fold

have been reported using GM technology (Trijatmiko et al., 2016).

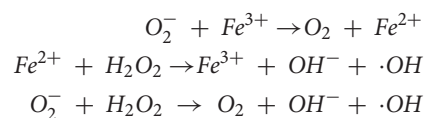
Even with this advantage however, the release of GM food crops is controversial due to public and political concerns for environmental and human safety. Many of such concerns are directed toward herbicide-resistant GM crops. Conversely, the current direction of GM focuses more on functional foods and is more subtle, leaning toward a cis-genic rather than transgenic approach. Recent years have also seen the rapid rise of genome editing techniques—these are more specific than GM, being capable of targeting specific genome locations for modification whilst also being integration-free. In genome editing, artificial nucleases are used for targeted gene integration or deletion. Four systems have developed—meganucleases, zinc finger nucleases (ZFN), transcription factor nucleases (TALENs), and clustered regularly interspaced short palindromic repeats (CRISPR-CAS) (Sander and Joung, 2014; Bortesi and Fischer, 2015). These have been tested in a wide range of species including crops like rice and wheat, though it has yet to be applied for biofortification purposes.

To date, starchy staples that contain little micronutrients like cereals, root crops, and banana have the primary targets for iron biofortification (Namanya, 2011; HarvestPlus, 2015a; Banana21, 2016). The advantages of such targets is that they form the bulk of local diets and given proper processing, have a long shelf-life, allowing for efficient delivery of the biofortified micronutrient over a large spatial and temporal scale. A wealth of information has been generated concerning these crops as a consequence of extensive focus. Biofortification works using GM in particular, have largely concentrated on major graminaceous crops like rice, wheat, and maize. In contrast, existing studies in non-graminaceous plants were conducted mainly in model species like tobacco or *Arabidopsis* for characterization purposes. While not as extensive, some work has also been conducted in crop species like banana (Matovu, 2016), cassava (Narayanan et al., 2015) lettuce (Goto et al., 2000), and soybean (Vasconcelos et al., 2006). Given the physiological differences between the non-graminaceous and graminaceous plants, it is difficult to extrapolate the effectiveness of iron biofortification approaches in the latter to the former. As it stands, there remains much to be explored in terms of iron biofortification of non-graminaceous crops.

IRON METABOLISM IN PLANTS

Prior to attempting any biofortification strategies, the significance of iron in the plants and the underlying mechanisms governing its metabolism must first be understood. Iron is the fourth most common element in the Earth's crust and can exist in a wide range of oxidation states, of which the most common are the ferrous (Fe^{2+}) and ferric (Fe^{3+}) forms. By virtue of its high redox potential, it forms a key component of biological processes involving electron exchange such as DNA synthesis, oxygen transport, cellular respiration, and photosynthesis, where it participates in the form of a cofactor in iron complexes. Examples of such complexes include hemoglobin, DNA helicases, and catalase.

For all its biological significance however, iron metabolic pathways can be summarized with the imagery of a precarious transfer of a radioactive material between containment facilities. Biologically, free iron may result from iron overload and/or insufficient sequestration capacity of the organism (Pietrangelo, 2003). Left alone, free Fe^{2+} catalyzes the formation of hydroxyl ($\cdot\text{OH}$) radicals through the Fenton reaction, and the process repeats when Fe^{2+} is regenerated from the resultant Fe^{3+} through reduction by the superoxide radical (O_2^-) (Haber and Weiss, 1934). The summation of this self-perpetuating reaction is known as the Haber-Weiss reaction:



Reactive oxygen species (ROS) generated as a consequence of this reaction can react with cellular components to cause oxidative damage (Kehrer, 2000; Aisen et al., 2001; Papanikolaou and Pantopoulos, 2005; Jeong and Guerinot, 2009; Kobayashi and Nishizawa, 2012); however on the other hand, they also serve as important signaling molecules and are an integral part of the stress response (Apel and Hirt, 2004). The fine line between cytotoxicity and biological function, and the intimate association between iron and ROS production, highlights the significance of proper regulation of iron metabolic pathways.

Given that iron nutrition and metabolism in plants has been extensively reviewed over the years (see Table 5), this review will provide only a brief overview of the topic. Iron metabolic pathways can be divided into three main processes: uptake, translocation and storage. Despite its abundance iron has poor solubility under aerobic conditions, particularly in high pH and calcareous soils, necessitating its solubilization before uptake can occur. This process is mostly accomplished via root exudates, the composition of which varies in response to the plant's physiological state and needs. In response to iron deficiency, the plant triggers the production of factors that directly or indirectly aid iron solubilization. Enhanced concentration of glutamate, ribitol, and glucose were observed in the root exudates of iron-deficient maize, which were suggested to attract and support siderophore-producing bacterial communities to aid iron solubilization (Carvalhais et al., 2011). Notable increases were also observed in the production of organic acids like malate and citrate, which increase the availability of iron through dissolution of insoluble iron compounds (Jones et al., 1996; Sánchez-Rodríguez et al., 2014).

In addition to the aforementioned means, different plant species have adopted specific approaches toward solubilize and acquire iron. These were first categorized as Strategy I and Strategy II by Römheld and Marschner (1986) and later studies have served to cement this grouping. As the topic of iron metabolism has been extensively reviewed over the years (see Thomine and Lanquar, 2011; Kobayashi and Nishizawa, 2012; Brumbarova et al., 2015), the following sub-sections will only provide a brief overview of the process.

Strategy I Uptake Mechanism

Strategy I is a reduction-based strategy used predominantly by non-graminaceous species, which includes all plants except grasses. Under iron deficiency, proton extrusion occurs through the action of H^+ -ATPases (HA), resulting in the acidification of the rhizosphere and reduction of insoluble iron (Rabotti and Zocchi, 1994; Dell'Orto et al., 2000; Santi et al., 2005; Santi and Schmidt, 2009). Phenolic compounds may also be secreted to facilitate iron uptake (Rodríguez-Celma et al., 2013). Iron and iron chelates are then reduced at the root surface by ferric-chelate reductase oxidase (FRO), which reduces Fe^{3+} chelates to Fe^{2+} by transferring electrons across the plasma membrane (Robinson et al., 1999; Waters et al., 2002). FRO is a family of membrane-bound metalloredutases that transfer electrons from cytosolic NADPH across membranes to electron-accepting substrates on the other side. In addition to facilitating acquisition of iron from the soil, this capability is also utilized in localizations where iron reduction is required for transport and/or assimilation, such as in the mesophyll (Brüggemann et al., 1993), reproductive tissues (Waters et al., 2002; Li et al., 2004a), and chloroplast membranes (Jeong and Connolly, 2009).

Following reduction at the root surface, the resulting $Fe(II)$ ions are absorbed across the plasma membrane via the iron-regulated transporter (IRT) (Eide et al., 1996; Vert et al., 2002). IRT is a member of the zinc-regulated transporter, iron-regulated transporter-like protein (ZIP) family that functions as membrane-bound uptake transporter for Zn and Fe (Lin et al., 2009). In *Arabidopsis* and tomato, IRT1 has been identified as responsible for uptake from the soil (Bereczky et al., 2003; Vert et al., 2009), with loss of function producing a severely stunted and chlorotic phenotype (Varotto et al., 2002; Vert et al., 2002). Co-regulated with IRT1 is the AtIRT2 homolog, which facilitates subcellular transport of iron and localizes to vesicle membranes instead of plasma membranes (Vert et al., 2009).

Both FRO and IRT activity is regulated in response to iron concentrations and increases in response to iron deficiency (Robinson et al., 1999; Connolly et al., 2003; Vert et al., 2003). Enhanced ferric reduction in particular, has been considered a hallmark indicator of iron deficiency (Römheld and Marschner, 1981; Higuchi et al., 1995) and increased capacity for FRO activity confers increased tolerance to low iron (Connolly et al., 2003; Peng et al., 2015).

Strategy II Uptake Mechanism

Unlike Strategy I, Strategy II is used by graminaceous plants (grasses) and revolves around the use of the mugineic acid (MA) family phytosiderophores in iron acquisition. The MA biosynthetic pathway starts with the conversion of three units of S-adenosylmethionine (SAM) into nicotianamine (NA) by nicotianamine synthase (NAS) (Higuchi et al., 1994, 1999). NA is converted to a 3' keto-intermediate by nicotianamine aminotransferase (NAAT), before being reduced to deoxymugeneic acid (DMA) by deoxymugeneic acid synthase (DMAS) (Kanazawa et al., 1994; Bashir et al., 2006). DMA can subsequently be converted to other MAs through a series of hydroxylations (Mori and Nishizawa, 1989; Ma and Nomoto, 1993), increasing levels of which improves the affinity for Fe^{3+}

and chelate stability under acidic conditions (von Wirén et al., 2000). Synthesized MAs are secreted into the soil through the phytosiderophore efflux transporter TOM1 (Nozoye et al., 2011) and the resulting ferric complexes are then taken up by the roots through specialized transporters like YELLOW STRIPE 1 (YS1) and YELLOW STRIPE 1-like (YSL) (Curie et al., 2001; Murata et al., 2006; Inoue et al., 2009; Kobayashi and Nishizawa, 2012). The resulting $Fe(III)$ -phytosiderophore complex is subsequently taken up via specialized transporters and transported throughout the plant (Kawai et al., 2001). Unlike the reduction-based approach used in Strategy I, phytosiderophore uptake is not limited by high pH, thereby conferring an advantage where such conditions are present (Römheld and Marschner, 1986).

Translocation and Storage

Following acquisition into the root symplast, iron is transported across the root to the vascular tissue and subsequently to the rest of the plant. The translocation process itself is a multi-step process, involving symplastic movement across the Casparian strip and to the desired site; the loading, unloading and transport through the vascular tissue; as well as remobilization from source tissue (Kim and Guerinot, 2007).

During transport, iron is maintained as a chelated complex with ascorbate, citrate or NA (Brown and Chaney, 1971; Stephan and Scholz, 1993; Pich et al., 1994; Grillet et al., 2014). With graminaceous plants, iron may also be complexed to DMA or MAs for transport (Koike et al., 2004; Ishimaru et al., 2010). Such complexes are pH-dependent, as are their interactions with other iron chelators (von Wirén et al., 1999). NA for instance, chelates both $Fe(II)$ and $Fe(III)$ at a higher pH but will preferentially bind for the former at pH 7.0. When bound to $Fe(III)$ at equilibrium, the NA complex dominates at pH 7.0–9.0 while the structurally similar DMA complex dominates at pH 3.0–6.0. Citrate removes iron from NA at pH 5.5; and it must be converted to $Fe(III)$ -citrate even if $Fe(II)$ is the major form in which Fe is loaded into xylem.

As with the uptake process, non-graminaceous plants seem to rely on reduction-based strategy while graminaceous plants utilize a chelation-based one in which $Fe(III)$ undergoes little or no change in redox state. The use of both strategies may be present in a single species, of which the only known example is rice (Ishimaru et al., 2006). This combination may represent an adaptation to the submerged conditions in which rice and its wild relatives grow, where iron is more readily available in ferrous than ferric form. Whether a similar occurrence may be found in other species remains to be seen, though orthologs of genes associated with Strategy I have also been found in other graminaceous species. An example of this is ZmIRT1 from maize, which was purported to be involved in both uptake and translocation, particularly to the seeds (Li et al., 2015). It should be noted that while differences between both strategies primarily affect the uptake process, the involvement of molecular components in the translocation process has further implications for the overall physiology of the plant.

Upon reaching the sink tissue, iron is reallocated as a cofactor in various complexes, or bound to the iron storage molecule ferritin and stored in the apoplastic space and vacuoles (Briat and

Lobréaux, 1997). Subsequent translocation and remobilization may occur in response to developmental and physiological needs, such as during iron deficiency (Waters and Troupe, 2012), seed filling (Hocking and Pate, 1977; Burton et al., 1998; Garnett and Graham, 2005), senescence (Shi et al., 2012; Maillard et al., 2015), and nodulation (Strozycki et al., 2007).

PULSES AS A VEHICLE FOR BIOFORTIFICATION

Aside from the starchy staples, another group of crops has been targeted for biofortification, albeit to a lesser degree. Pulses, as defined by the FAO (1994), are leguminous crops harvested solely for dry grain. While this term encompasses most grain legumes, soybean and peanut are excluded from this classification, because they are traditionally viewed as oilseed crops (Pulse Australia, 2016b).

Like cereals, pulses have a long history of cultivation and have been a significant constituent in human diets since around 10,000 BC (Fuller et al., 2001; Caracuta et al., 2015). As a crop, pulses present two main benefits, both of which are complementary to cereals. The first is their agronomic characteristics. By virtue of their nitrogen fixing properties, pulses are often grown as an intercrop or as a mixed crop to replenish soil nitrogen levels, thereby reducing the need for fertilizers. Cultivation with pulse crops have also been shown to increase the uptake of nitrogen, sulfur, and phosphorus by cereals, resulting in an enhanced yield and grain quality (Li et al., 2003, 2004b; Agegnehu et al., 2006; Banik et al., 2006; Gooding et al., 2007). Yield stability is also increased (Rao and Willey, 1980).

The second benefit of pulses is their nutritional density. Pulses are a rich source of carbohydrates and fiber. Their most prominent feature however, is their high protein content of 21–26% and an amino acid profile complementary to that of cereals, being rich in lysine, leucine, and arginine (Phillips, 1993; Iqbal et al., 2006; Pulse Canada, 2016). Their excellence as a vegetarian source of protein and affordability in contrast to livestock products has earned them the famous moniker of “poor man’s meat.” Pulses are also rich in micronutrients like folate, thiamine, riboflavin, niacin, calcium, magnesium, iron, and zinc (Phillips, 1993; Iqbal et al., 2006; Jukanti et al., 2012). Other than contributing to macro- and micronutritional needs, several health benefits have been associated with inclusion of pulses in the diet. Their low glycemic index (GI) has been linked to the management of diabetes and diabetes-related diseases (Rizkalla et al., 2002; Sevenpiper et al., 2009) while bioactive components have been investigated for their health potential—e.g., lectins for their immunomodulatory effect, protease inhibitors for anti-inflammatory effects, and angiotensin I-converting enzyme (ACE) inhibitory peptides for their anti-hypertensive properties (Rochfort and Panozzo, 2007; Roy et al., 2010).

Despite their agronomic and nutritional benefits, pulses have not received the same amount of attention or development as the main starchy staples. Between 1961 and 2014, pulse yield and production values increased by 42.3 and 90.4%, respectively, a small fraction compared to the increase of 187.2 and 219.4% in

cereals (FAO, 2016b). Much of this disparity can be attributed to developments made during the Green Revolution, in which the focus on productivity and protein-calorie malnutrition led to the shift from cultivation of traditional micronutrient-rich crops to the more productive and profitable starchy cereals (Pinstrup-Andersen and Hazell, 1985; Pingali, 2012). Poor policy and diversion of land to cereal cultivation has led to a reduction in pulse supply, effectively driving prices up and decreasing consumption per capita (Kennedy and Bouis, 1993; Kataki, 2002; Akibode and Maredia, 2012).

As highlighted in the special feature on pulses in the 2014 Food Outlook (FAO, 2014), recent years have seen several key changes in pulse production and trade. Asia remains the region with the highest pulse production, with India continuing as the largest pulse producing country, contributing at least 20% toward global pulse production (Figure 1 and Table 4). Production in other regions except Europe has also increased, fueled by domestic and international demand. In contrast to these countries is China, whose production has decreased due to a number of factors such as population increase and decreasing availability of arable land. Despite the shift in preference for animal-based products and protein that accompanies growing affluence, India and China remain major importers, consuming approximately 40% of the world’s pulse production as food, and 30% as feed. Much of this is provided by major exporters like Canada and Australia. With other major producers like Myanmar and Brazil, pulse consumption is primarily domestic.

Pulse production, consumption, and trade are expected to increase alongside population growth, particularly with increasing promotion from government campaigns (Akibode and Maredia, 2012) and the declaration of 2016 as the “International Year of Pulses” by the UN General Assembly. Increasing awareness and concern over nutritional composition of food, particularly by food manufacturers, has attracted greater interest in pulses, which will likely translate into further support and development of the pulse industry (FAO, 2014).

CHALLENGES OF PULSE BIOFORTIFICATION

Concerning the biofortification of pulses however, there are some challenges. While pulses are a diverse group featuring a wide variety of species and cultivars, this has presented itself as a double-edged sword. The genetic richness is a treasure trove that lends itself to crop improvement, but has simultaneously resulted in a lack of a concerted global effort at development. Development has also thus far has been focused on yield, disease tolerance, and macronutrient quality, with little to no emphasis on other nutritional aspects. However, this has changed over the past decade with the growing interest in micronutrient content. Increasing numbers of genotypes and cultivars being assayed for their iron and zinc composition (Ray et al., 2014; Thavarajah et al., 2014; Santos and Boiteux, 2015; Blair et al., 2016). Concerning outputs however, there has been little published work on pulse biofortification, with even fewer examples of biofortified pulses. Currently the only known example of a

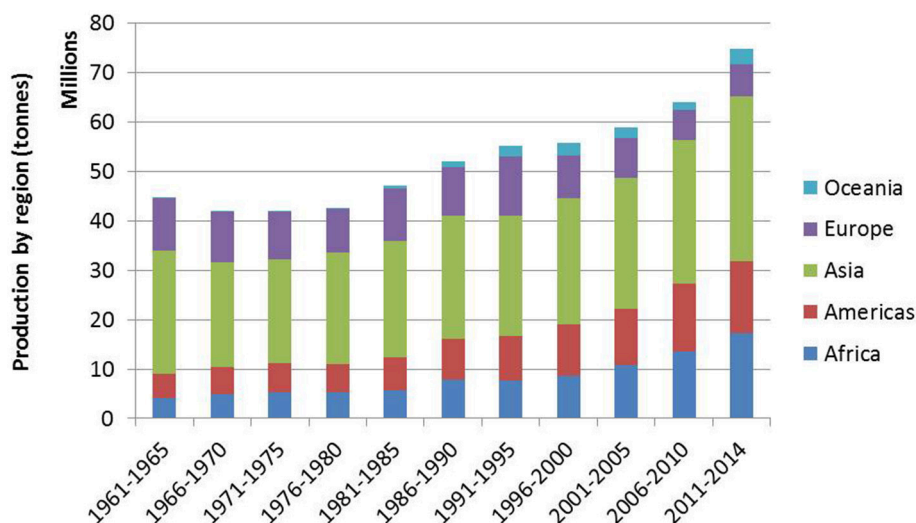


FIGURE 1 | Average pulse production by region (FAO, 2016a).

biofortified pulse is the high iron common bean generated from the HarvestPlus breeding program; to date, several varieties have been produced with improvements in iron content ranging from 47 to 94% (Katsvairo, 2015).

Another challenge with pulse biofortification is the bioavailability of iron. Bioavailability, as defined by Carpenter and Mahoney (1992), is the “proportion of a nutrient present in food that the body is able to absorb and utilize by incorporation into physiologically functional pools.” It is a complex and multifaceted issue subject to a range of physiological and physiochemical factors such as those illustrated in **Table 6**. In humans, iron uptake occurs in the duodenum and upper jejunum in the intestinal tract. Regulation is done via absorption rather than excretion, and occurs in response to iron status and rate of erythropoiesis (Wheby et al., 1964; Bothwell, 1995). Under normal conditions, absorption is inversely correlated to the state of the body’s iron stores with increased uptake by the intestinal mucosal cells as iron stores are depleted (Walters et al., 1975).

For uptake to occur however, the mineral must be in a form absorbable by the mucosal cells. Dietary iron can be classified as heme or non-heme based on its attachment to heme proteins or lack thereof. Both form separate pools in the gastrointestinal tract and undergo different absorption pathways (Björn-Rasmussen et al., 1974). Compared to its heme counterpart, non-heme iron is the predominant form found in plants and has a lower bioavailability and is particularly susceptible to the influence of dietary factors (Björn-Rasmussen et al., 1974). This is of especial significance in pulses due to the abundance of naturally occurring inhibitors which bind to iron and prevent uptake. Examples of such inhibitors include phytic acid, polyphenols, tannins, and fiber (Sandberg, 2002; Ghavidel and Prakash, 2007; Thavarajah et al., 2014).

Phytic acid (PA), also known as inositol hexaphosphate or IP6, has been identified as a one of the major inhibitors of iron

bioavailability. PA serves as the principal form of phosphorus storage in seeds, where it is present as a phytate salt of mineral cations like potassium, magnesium, calcium, manganese, and zinc. Depending on the species, cultivar and conditions of growth, it may constitute 40–84% of total seed phosphorus (Lolas et al., 1976; Griffiths and Thomas, 1981; Ravindran et al., 1994). Amongst the inositol polyphosphates, the lower inositol phosphates IP2, IP3, and IP4 play a minor role in inhibiting iron bioavailability (Sandberg et al., 1989, 1999). The main inhibitors are IP6 and IP5, which are capable of reducing iron solubility by 38.8 and 33%, respectively, through the formation of insoluble complexes (Sandberg et al., 1989).

The impact of PA on biofortification efforts can be illustrated using the example of the biofortified beans. Despite their higher iron content, feeding trials conducted in Rwanda have indicated iron bioavailability of biofortified beans was similar, if not lower, than that of the unfortified beans (Petry et al., 2012, 2014). As a result, while the total of iron absorbed from the biofortified beans was higher than the unfortified beans, it was considerably less than expected. As a means to improve the effectiveness of biofortification, the reduction in PA concentration was recommended by the authors (Petry et al., 2012, 2014).

This recommendation has been applied in several cereal crops (Larson et al., 1998, 2000; Raboy et al., 2000; Guttieri et al., 2004) and more recently in bean (Campion et al., 2009). The effectiveness of the low-phytic acid bean lines is currently inconclusive however, as bioavailability assessments have yielded conflicting results due to differences in experimental design (Petry et al., 2013, 2016). Poor cooking quality was also observed in the low-phytic acid seeds, which may have contributed the adverse gastrointestinal side-effects in the participants in one of the studies (Petry et al., 2016). The relationship between phytic acid and cooking quality have been alluded to in other studies on lentil and bean (Kon and Sanchuck, 1981; Bhatti and Slinkard, 1989). Interestingly, no such effect was reported in low-phytic

TABLE 4 | Growth in production of major pulse producers (FAO, 2014, 2016b).

Country	Production (mmt)				Principal pulse crops
	1961	1981	2001	2014	
India	12.9	10.8	12.2	19.98	Chickpeas, beans, pigeon peas
Myanmar	0.2	0.4	2.0	5.0	Beans, pigeon peas, chickpeas
Canada	0.1	0.2	3.4	5.8	Peas, lentils
China	8.5	6.4	5.1	4.5	Beans, broad beans, peas
Brazil	1.8	2.4	2.5	3.3	Beans
Nigeria	5	0.6	2.3	2.2	Cowpeas
Ethiopia	0.6	0.9	1.2	2.6	Broad beans, beans, chickpeas, peas
Australia	0	0.3	2.7	3.0	Lupines, lentils, chickpeas
USA	1.1	1.7	1.3	2.4	Beans, peas
Tanzania, U. Rep.	0.1	0.3	0.8	1.8	Beans
Rest of the world	15.0	17.5	22.6	27.02	
Total	40.8	41.6	55.9	77.6	

acid maize lines (Mendoza et al., 1998); whether this is a legume-specific issue remains to be confirmed. Aside from influencing cooking quality, phytic acid is also known to have antioxidant properties and protective effects against heart disease and cancer (Sharma, 1986; Nelson et al., 1988; Vucenik and Shamsuddin, 2003). It is unknown if reduction in phytic acid content would affect such properties. Similarly, the subsequent long-term effect on human health is unknown.

For all the challenges presented in this section however, there is much potential to be explored in pulse biofortification. Much of the existing knowledge concerning this area is limited to the work done on the iron biofortified bean. That has proven to be a successful means of alleviating iron deficiency, promising much for other pulses.

CHICKPEA AS A TARGET FOR BIOFORTIFICATION

Chickpea (*Cicer arietinum*) is an important pulse crop that has been cultivated by humans since the Stone Age. As of 2009, it is the second most important pulse crop in the world after the common bean, having overtaken peas as the pulse crop with the second highest global production values. Global production has climbed steadily since 2008 to exceed 14.2 million tons in 2014, of which approximately 96% is grown in developing countries (FAO, 2016b). India in particular, has historically been the largest producer and consumer of chickpea; in 2013 alone it contributed approximately 65 and 33% to total chickpea production and import, respectively (FAO, 2016b). In terms of consumption, it is difficult to obtain precise statistics due to the lack of available data. However, based on calculations using production and trade values, the global average for chickpea consumption was estimated to be around 1.3 kg/year per person between 2006 and 2008, with South Asia and the Middle East-North Africa

TABLE 5 | List of general reviews on iron nutrition and metabolism.

- Hell and Stephan, 2003
- Kim and Guerinot, 2007
- Walker and Connolly, 2008
- Morrissey and Guerinot, 2009
- Thomine and Lanquar, 2011
- Kobayashi and Nishizawa, 2012
- Thomine and Vert, 2013
- Briat et al., 2015
- Brumbarova et al., 2015
- Connorton et al., 2017
- Curie and Mari, 2017

regions being the biggest consumers at 4.25 kg/person and 2.11 kg/person per year, respectively (Akibode and Maredia, 2012). The demand is predicted to grow, particularly in Africa and Asia, due to population increase and increasing support from the governments in encouraging pulse consumption (Rao et al., 2010; Akibode and Maredia, 2012). This increase in demand is not limited to those regions; in the USA for instance, net domestic use of chickpea nearly doubled from 199.6 g in 2010 to 322.1 g in 2014 (Wells, 2016).

Most of the chickpea in the global market can be classified into two main types which are primarily distinguishable by their seed morphology, specific aspects of which influence their end-use. The first type is the kabuli, also known as garbanzos. Kabuli seeds are large and round, weighing approximately 400 mg per seed (Pulse Australia, 2016a). The seed coat is thin and light-colored, ranging from shades of white to cream and the seeds are typically consumed whole or made into hummus (Gaur et al., 2015; Pulse Australia, 2016a). Kabuli cultivation areas are mostly located in Southern Europe, Northern Africa, Afghanistan, Pakistan, Chile, and India (Gaur et al., 2015).

The second type is the desi, which forms the bulk of the international export market (Rao et al., 2010). Desi seeds are small, wrinkled, and angular, with an approximate weight of 120 mg per seed (Pulse Australia, 2016a). The seed coat is also 1.2 to 3 times thicker than the kabuli (Umaid et al., 1984; Wood et al., 2011) and can be found in a greater variety of colors ranging from brown to yellow, as well as orange, black and green. Desi seeds are commonly dehulled and split to obtain the cotyledons, which are then known as chana dhal and can in turn be milled to flour, known as besan or gram flour.

As a food crop, chickpea can be utilized in a variety of ways. Green pods, immature seeds and young leaves can be consumed as a vegetable while the stover and pod husks can be used as animal feed (Ibrikci et al., 2003; Yadav et al., 2007). The primary commodity however, is the dried mature seed which can be used as animal feed or for human consumption. With the latter, the long history of consumption in various regions such as India, the Middle East, and Europe has given rise to a diversity of dishes in which chickpea can be utilized. Chickpeas are consumed on their own or with other foods; seeds may be eaten whole, hulled, or ground into flour from which other products may be derived. Preparation for consumption can be by various processing methods such as soaking, sprouting,

TABLE 6 | Factors affecting bioavailability of some trace elements (House, 1999).

Host factors	Dietary factors	
	Dietary composition	Food preparation
<ul style="list-style-type: none"> • Age • Sex • Ethnic background <ul style="list-style-type: none"> ◦ Types of food selected ◦ Geographic living area • Economic status <ul style="list-style-type: none"> ◦ Type, quality, and quantity of selected food • Physiological status <ul style="list-style-type: none"> ◦ Pregnancy ◦ Lactation ◦ Physical activity • Nutritional status <ul style="list-style-type: none"> ◦ Moderate or frank deficiency ◦ Lean body mass • Disease (including parasitism) 	<ul style="list-style-type: none"> • Protein quality <ul style="list-style-type: none"> ◦ Protein source ◦ Animal vs. plant protein ◦ Amino acid balance • Protein quantity • Trace element quantity • Physiochemical form of trace element • Nutrient interactions <ul style="list-style-type: none"> ◦ Element–element ◦ Element–organic compounds • Promoters <ul style="list-style-type: none"> ◦ Meat ◦ Ascorbate ◦ Citrate ◦ Vitamin D ◦ Some amino acids ◦ Some sugars • Inhibitors <ul style="list-style-type: none"> ◦ Phytate ◦ Oxalate ◦ Polyphenols ◦ Fiber ◦ Goitrogens ◦ Excess ascorbate or folate • Micronutrient deficiencies <ul style="list-style-type: none"> ◦ Ascorbate ◦ Riboflavin ◦ Vitamin E 	<ul style="list-style-type: none"> • Raw • Cooking (various methods) • Fermentation • Malting • Milling • Extraction • Soaking

fermenting, boiling, steaming, roasting, extrusion, and puffing (Yadav et al., 2007), all of which exert different effects on the overall nutritional quality (Poltronieri et al., 2000; Sebastián et al., 2001; Ghavidel and Prakash, 2007; Hemalatha et al., 2007a).

Much like other pulses, the nutritional qualities of chickpea have long been recognized and documented. In addition to high protein content (20–22%), chickpeas are also rich in micronutrients like folate, magnesium, zinc, and iron (USDA, 2013). Studies conducted by different authors have found iron content to range from 2.4 to 11 mg/100 g (e.g., USDA¹; Meiners et al., 1976; Wood and Grusak, 2007). Likewise, various studies have reported differing values for phytic acid and other antinutrients (e.g., Chitra et al., 1995; Ghavidel and Prakash, 2007; Hemalatha et al., 2007b), indicating a possible effect of genotype and environmental factors on overall iron bioavailability. When measured as dialyzable iron generated from a simulated gastrointestinal digest, bioavailability has been found to vary widely across different studies, ranging from about 6 to 25% (Chitra et al., 1997; Ghavidel and Prakash, 2007; Hemalatha et al., 2007b). The reason behind this disparity is as yet unclear,

¹USDA Basic Report: 16056, *Chickpeas (Garbanzo Beans, Bengal Gram), Mature Seeds, Raw*.

though analytical procedures and variations in samples have been suggested as a possible cause (Platel and Srinivasan, 2016). Given the multifaceted nature of nutrient bioavailability, the values obtained are at best relative.

Regardless of processing methods and culinary adjustments, the composition of the starting material is vital. As demonstrated by Petry et al. (2012), enhancement of iron content alone does not necessarily translate into an improved iron status of the consumer, particularly when there is a concurrent enhancement in phytate content. Both the iron content and overall composition of the grain itself should therefore be considered in biofortification strategies. However, in light of the lack of information concerning bioavailability, it would be prudent for biofortification efforts to first target total seed iron content before progressing to bioavailability. Considerable progress has been made to that end, particularly with the growing interest in chickpea as a target for iron biofortification. While a concerted global effort has yet to materialize, pockets of development have emerged with India and Canada at the forefront. To date, the chickpea genome has been sequenced (Varshney et al., 2013). Chickpea populations in those countries have also been screened for genetic diversity and iron accumulation traits, allowing for identification of the associated QTLs (Diapari et al., 2014; Upadhyaya et al., 2016). In terms of biofortification via GM, no work has been done yet. It is however, a viable option—while chickpea can be considered a recalcitrant species, successful transformation protocols have been established (Sarmah et al., 2004; Indurker et al., 2010). Such work would also provide insight into the physiological workings; this in turn can inform later biofortification efforts by identifying specific traits or mechanisms which can be targeted by breeding or GM.

Given the relative youth of this endeavor to biofortify chickpea for iron, no biofortification targets have yet been set. As stated by Bouis and Welch (2010), several variables need be considered in the setting of such targets. The challenge lies primarily in the lack of information concerning the different variables in chickpea. Unlike the common bean which serves as a staple, chickpea is a secondary staple and depending on the type and cultivar, may be processed into various forms for consumption (Yadav et al., 2007). This would in turn affect iron content and bioavailability. Consequently, the consumption profile for chickpea is expected to be lower and potentially more varied compared to the common bean, particularly across different age and cultural demographics. Until more detailed and specific information concerning chickpea is obtained, only general assumptions may be made. In the interim however, efforts can be made to understand and engineer for increased iron content.

POTENTIAL APPROACHES TO ENGINEERING FOR ENHANCED IRON CONTENT IN CHICKPEA

In the interest of iron biofortification, five rate-limiting steps to grain iron accumulation have been identified by Sperotto et al. (2012): (1) uptake from soil, (2) xylem loading in roots, (3)

phloem transport from leaves, (4) unloading for grain filling, and (5) grain sink strength. These steps can be classified into the three main processes of uptake, translocation, and storage. Genes associated with these processes have been identified as promising candidates for iron biofortification, and over the past decade several of them have been applied to different plant species.

Rice—A Case Study

Amongst these, rice can be considered the flagship for transgenic iron biofortification. Since rice is a grain crop, the lessons learnt may, in part, be transferrable to chickpea. As illustrated in the review by Masuda et al. (2013a), several gene combinations targeting the uptake, translocation, storage or any combination of the three processes have been attempted to differing levels of success. Of the combinations investigated thus far, those containing NAS and ferritin (FER) have yielded the most promising results. Individually, NAS and FER have been demonstrated to enhance iron accumulation. Overexpression of the former for instance, increased iron content in polished grains by four-fold (Johnson et al., 2011), while overexpression of a soy homolog of the latter produced up to a 3.7-fold increase (Vasconcelos et al., 2003). Similar results have been obtained when combined with other genes involved in iron transport or MAs synthesis. Approximately three-fold increase was obtained from overexpression of the rice yellow stripe like-2 (YSL2), barley NAS1, and soybean FER genes (Masuda et al., 2012). A four-fold increase was obtained from overexpressing soybean FER in conjunction with barley NAS1, two NAAT genes and a mugineic acid synthase gene (Masuda et al., 2013b).

The best results however, were achieved simply by using just the NAS-FER combination. Constitutive expression of *Arabidopsis* NAS1 together with barley FER and a fungal phytase, both under the regulation of a rice seed storage globulin promoter, enhanced iron accumulation in rice endosperm by up to six times (Wirth et al., 2009). More recently, up to 7.5-fold increase in iron content was obtained from transgenic rice through constitutive overexpression of OsNAS2 and seed-specific expression of soybean FER (Trijatmiko et al., 2016). The enhancement was not limited to total iron content alone, as iron bioavailability was similarly improved. This was demonstrated in studies with Caco-2 cell cultures, where the amount of iron absorbed from the transgenic lines was more than double that of the controls, even in the absence of any bioavailability enhancers (Trijatmiko et al., 2016).

Tailoring a GM Approach to Chickpea

Whether a similar feat may be emulated in other species is yet unknown as the NAS-FER gene combination has only been applied in rice. In this case, the success of the NAS-FER gene combination may be attributed to the unique role of NA in the non-graminaceous system. In it, NA, by virtue of its role in the synthesis of DMA and MAs, is directly involved in the uptake and translocation processes (Wang et al., 2013). Through combination with FER, it allows for the simultaneous

targeting of all three major processes of iron metabolism, thereby overcoming the rate-limiting steps listed by Sperotto et al. (2012). Incidentally, the issue of bioavailability is also resolved—NA is a known enhancer of iron bioavailability (Zheng et al., 2010), while ferritin has a bioavailability equivalent to that of ferrous sulfate which is used in iron supplements (Davila-Hicks et al., 2004; Lönnerdal et al., 2006).

Based on these observations, there is reason to believe that application of the NAS-FER approach to other graminaceous species will yield similar outcomes to rice. The effect however, may be limited in non-graminaceous crops like chickpea due to the absence of the MAs biosynthetic pathway, which diminishes the contribution of NAS to the uptake process. Such is evident when comparing the results of NAS overexpression in graminaceous and non-graminaceous species, where greater enhancement in iron content was observed in the former (Masuda et al., 2009; Johnson et al., 2011) than the latter (Douchkov et al., 2005; Cassin et al., 2009). In any case, prior studies in model species like *Arabidopsis* and tobacco have confirmed the individual effect of each gene (Van Wuytswinkel et al., 1999; Douchkov et al., 2005; Cassin et al., 2009); should the NAS-FER approach be applied to chickpea, some enhancement of seed iron content can still be expected. Similarly, iron bioavailability can increase, though the extent is difficult to predict given the higher levels of inhibitors in chickpeas compared to rice (Hemalatha et al., 2007b).

At the moment these are speculations—with the existing transgenic research concentrated on cereals, there is little precedent for reliable extrapolation to a leguminous crop. Nonetheless, two main principles can be drawn from the success of the NAS-FER strategy in rice, and that is (1) the simultaneous targeting of multiple rate-limiting steps, and 2) targeting of bioavailability in addition to iron content.

Targeting Multiple Rate-Limiting Steps

One of the main challenges to engineering for iron accumulation in chickpea is the lack of specific knowledge on iron homeostasis in chickpea. Even amongst closely related members of non-Gramineae, interspecies variation exists as different mechanisms or components may be favored. Such is evident when comparing studies on QTLs and iron accumulation traits—differing suites of associated genes have been found in soybean (Ning et al., 2015) and chickpea (Upadhyaya et al., 2016).

Examination of such genes may yield potential candidates for use in GM biofortification. Drawing from the example of NAS in rice, selection of such candidate genes can be based on their involvement in both uptake and translocation processes, though extra emphasis should be placed on the former. As mentioned by Sperotto et al. (2012), the ability to access soil iron under differing environmental conditions is the first major bottleneck to iron accumulation in plants, with poor uptake capacity typically associated with susceptibility to iron deficiency (Mahmoudi et al., 2007; Waters and Troupe, 2012).

Several genes fitting such criteria can be found in the list by Upadhyaya et al. (2016), and key examples include transporters like IRT, FRO, YSL, NRAMP (natural resistance-associated macrophage protein), and zinc-regulated transporter,

iron-regulated transporter-like protein (ZIP). As these are involved in both uptake and translocation (Vert et al., 2002; Lanquar et al., 2005; Vasconcelos et al., 2006), targeting them may, theoretically, allow for two processes to be simultaneously enhanced. Constitutive overexpression of AtFRO2 in soybean for instance, was found to increase both iron uptake and leaf iron content (Vasconcelos et al., 2006). However, this effect may be dependent on the homolog used as each may serve specific functions. Within FRO family in *Arabidopsis* for instance, AtFRO2 is expressed in the roots and facilitates uptake during iron deficiency (Connolly et al., 2003), while FRO7 is expressed in the chloroplasts where it contributes to iron supply (Jeong et al., 2008).

A potential pitfall to such an approach is the inherent regulatory mechanisms. Unlike NAS, which appears to be amenable to manipulation with little to no side-effects (Pianelli et al., 2005; Johnson et al., 2011; Lee et al., 2012), transporters like IRT and FRO appear to be regulated by mechanisms which are less forgiving to interference. In *Arabidopsis* overexpressing AtIRT1 or AtFRO2, no increase in root reduction or protein levels was observed under iron-sufficient conditions due to post-transcriptional regulation (Connolly et al., 2002, 2003). However, no such impediment was noted when the maize homolog, ZmIRT1, was overexpressed, with transgenic lines having significantly higher seed iron contents compared to the wild-type (Li et al., 2015). This discrepancy was attributed to the low homology between the ZmIRT1 and the native IRT genes, which may in turn point to a means of bypassing post-transcriptional regulation.

Aside from enhancing uptake and translocation, sink strength may also be targeted. As far as GM biofortification efforts go, this has traditionally been done using FER. The application in chickpea appears to be highly feasible as amongst the fifteen genes associated with seed iron accumulation in chickpea, two were identified as ferritins. Strong constitutive FER overexpression however, carries the risk of excessive iron sequestration, resulting in manifestation iron deficiency symptoms (Van Wuytswinkel et al., 1999). This may be avoided through seed-specific expression, particularly in the cotyledons which are the main products. Much like the NAS-FER approach in rice, a multigenic approach combining FER with one of the aforementioned transporters may also be used. This will also allow for simultaneous targeting of all three major processes of iron metabolism, thereby overcoming the bottlenecks described by Sperotto et al. (2012). Theoretically, such a multigenic approach may also translate to higher levels of iron accumulation through synergy between the transgenes. Such was observed in rice, where the NAS-FER combination produced a synergistic effect (Wirth et al., 2009; Trijatmiko et al., 2016), resulting in higher seed iron contents compared to the monogenic NAS approach (Masuda et al., 2009; Johnson et al., 2011; Lee et al., 2011).

Targeting Bioavailability

As previously mentioned, the use of FER has the added benefit of enhancing iron bioavailability in addition to total iron content. Concerning the enhancement of bioavailability however, the use of FER is but one means. Other options may include targeting the

concentrations of inhibitors or enhancers. The former has already been attempted in maize and rice through the overexpression of phytase, and increases in bioavailability have been reported (Lucca et al., 2001; Drakakaki et al., 2005; Wirth et al., 2009). While promising, the actual effect on human health is unknown as the assays were done using *in vitro* methods. Given that no negative consequences were observed from the addition of exogenous phytase to food (Hurrell et al., 2003), it is likely that the detrimental effects observed with low-phytate beans (Petty et al., 2016) may be avoided.

Concerning bioavailability enhancers, some work has already been done in the form of NAS-overexpressing crops and the results discussed in the above sections (Johnson et al., 2011; Trijatmiko et al., 2016). An alternative candidate is ascorbic acid, a potent enhancer occurring naturally in plants which has been demonstrated to prevent the inhibitory effects of phytate and polyphenols (Hallberg et al., 1989; Siegenberg et al., 1991). However, while promising, ascorbic acid is also infamous for its thermal instability (Van den Broeck et al., 1998; Munyaka et al., 2010), with cooking generally resulting in degradation (Sood and Malhotra, 2002; Moriyama and Oba, 2008). Its effectiveness in a transgenic biofortification strategy is therefore questionable given the processing requirements of a grain crop like chickpea.

SUMMARY AND IMPLICATIONS

In summary, iron deficiency is a global health problem which may be alleviated through the use of biofortified crops. Transgenic biofortification efforts, as well as most studies on iron metabolism, thus far have largely been directed at cereal crops like rice. As members of the Gramineae family, their molecular biology and physiology differ significantly from their non-graminaceous counterparts. Consequently, biofortification strategies successfully applied in a graminaceous species like rice may behave differently in a non-graminaceous species. The extent of this difference is currently unclear as studies have primarily been performed in model species like *Arabidopsis* and tobacco primarily for gene characterization. There is a need to tailor specific biofortification strategies for use in non-graminaceous species. This is particularly so for important secondary staples like pulses—population growth as well as environmental pressures has increased the demand for affordable, water-efficient sources of protein. As the second most important pulse crop in the world, chickpea stands in a unique position to meet this need. It is widely consumed in the Asian and African regions where population growth, as well as the incidence of iron deficiency, is highest. The iron biofortification of chickpea can therefore serve as a sustainable means to alleviate the public health burden where it is heaviest.

While some breeding work is currently underway, there has been no recorded attempts to biofortify chickpea via a GM approach. However, the avenue to do so is available, given the establishment of successful transformation protocols. Valuable lessons can be learnt from the success of the GM biofortified rice and applied to the formulation of biofortification strategies for pulse crops like chickpea. Existing QTL and trait analysis have

identified several candidate genes which may be used to enhance iron content and/or bioavailability, opening up new doors for further exploration.

AUTHOR CONTRIBUTIONS

This review was part of a larger project designed and headed by SM, BW, AJ, and SD. The document was written by GT, TMLH

and MRK. Drafts were edited by the other authors, and upon their approval, was submitted for publication.

FUNDING

This review was written as part of the Tropical Pulses for Queensland, funded by the Queensland Government.

REFERENCES

- Aciksoz, S. B., Yazici, A., Ozturk, L., and Cakmak, I. (2011). Biofortification of wheat with iron through soil and foliar application of nitrogen and iron fertilizers. *Plant Soil* 349, 215–225. doi: 10.1007/s11104-011-0863-2
- Agegehu, G., Ghizaw, A., and Sinebo, W. (2006). Yield performance and land-use efficiency of barley and faba bean mixed cropping in Ethiopian highlands. *Eur. J. Agron.* 25, 202–207. doi: 10.1016/j.eja.2006.05.002
- Aisen, P., Enns, C., and Wessling-Resnick, M. (2001). Chemistry and biology of eukaryotic iron metabolism. *Int. J. Biochem. Cell Biol.* 33, 940–959. doi: 10.1016/S1357-2725(01)00063-2
- Akibode, S., and Maredia, M. K. (2012). *Global and Regional Trends in Production, Trade and Consumption of Food Legume Crops*. Michigan State University, Michigan, USA.
- Alvarez-Uria, G., Naik, P. K., Midde, M., Yalla, P. S., and Pakam, R. (2014). Prevalence and severity of anaemia stratified by age and gender in rural India. *Anemia* 2014:176182. doi: 10.1155/2014/176182
- Anuradha, K., Agarwal, S., Rao, Y. V., Rao, K. V., Viraktamath, B. C., and Sarla, N. (2012). Mapping QTLs and candidate genes for iron and zinc concentrations in unpolished rice of Madhukar × Swarna RILs. *Gene* 508, 233–240. doi: 10.1016/j.gene.2012.07.054
- Apel, K., and Hirt, H. (2004). Reactive oxygen species: metabolism, oxidative stress, and signal transduction. *Annu. Rev. Plant Biol.* 55, 373–399. doi: 10.1146/annurev.arplant.55.031903.141701
- Baltussen, R., Knai, C., and Sharan, M. (2004). Iron fortification and iron supplementation are cost-effective interventions to reduce iron deficiency in four subregions of the world. *J. Nutr.* 134, 2678–2684.
- Banana21 (2016). *Biofortification*.
- Banik, P., Midya, A., Sarkar, B. K., and Ghose, S. S. (2006). Wheat and chickpea intercropping systems in an additive series experiment: advantages and weed smothering. *Eur. J. Agron.* 24, 325–332. doi: 10.1016/j.eja.2005.10.010
- Bashir, K., Inoue, H., Nagasaka, S., Takahashi, M., Nakanishi, H., Mori, S., et al. (2006). Cloning and characterization of deoxymugineic acid synthase genes from graminaceous plants. *J. Biol. Chem.* 281, 32395–32402. doi: 10.1074/jbc.M604133200
- Batool, R., Butt, M. S., Sultan, M. T., Saeed, F., and Naz, R. (2013). Protein-energy malnutrition; a risk factor for various ailments. *Crit. Rev. Food Sci. Nutr.* 55, 242–253. doi: 10.1080/10408398.2011.651543
- Bereczky, Z., Wang, H.-Y., Schubert, V., Ganai, M., and Bauer, P. (2003). Differential regulation of *nramp* and *irt* metal transporter genes in wild type and iron uptake mutants of tomato. *J. Biol. Chem.* 278, 24697–24704. doi: 10.1074/jbc.M301365200
- Bhatty, R. S., and Slinkard, A. E. (1989). Relationship between phytic acid and cooking quality in lentil. *Can. Inst. Food Sci. Technol. J.* 22, 137–142. doi: 10.1016/S0315-5463(89)70349-7
- Björn-Rasmussen, E., Hallberg, L., Isaksson, B., and Arvidsson, B. (1974). Food iron absorption in man. Applications of the two-pool extrinsic tag method to measure heme and nonheme iron absorption from the whole diet. *J. Clin. Invest.* 53:247. doi: 10.1172/JCI107545
- Blair, M. W., Astudillo, C., Grusak, M. A., Graham, R., and Beebe, S. E. (2009). Inheritance of seed iron and zinc concentrations in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* 23, 197–207. doi: 10.1007/s11032-008-9225-z
- Blair, M. W., Knewton, S. J., Astudillo, C., Li, C.-M., Fernandez, A. C., and Grusak, M. A. (2010). Variation and inheritance of iron reductase activity in the roots of common bean (*Phaseolus vulgaris* L.) and association with seed iron accumulation QTL. *BMC Plant Biol.* 10:215. doi: 10.1186/1471-2229-10-215
- Blair, M. W., Wu, X., Bhandari, D., and Astudillo, C. (2016). Genetic dissection of ICP-detected nutrient accumulation in the whole seed of common bean (*Phaseolus vulgaris* L.). *Front. Plant Sci.* 7:219. doi: 10.3389/fpls.2016.00219
- Bortesi, L., and Fischer, R. (2015). The CRISPR/Cas9 system for plant genome editing and beyond. *Biotechnol. Adv.* 33, 41–52. doi: 10.1016/j.biotechadv.2014.12.006
- Bothwell, T. H. (1995). Overview and mechanisms of iron regulation. *Nutr. Rev.* 53, 237–245. doi: 10.1111/j.1753-4887.1995.tb05480.x
- Bouis, H. E., and Welch, R. M. (2010). Biofortification — a sustainable agricultural strategy for reducing micronutrient malnutrition in the global south. *Crop Sci.* 50, S20–S32. doi: 10.2135/cropsci2009.09.0531
- Briat, J.-F., and Lobréaux, S. (1997). Iron transport and storage in plants. *Trends Plant Sci.* 2, 187–193. doi: 10.1016/S1360-1385(97)85225-9
- Briat, J.-F., Dubos, C., and Gaymard, F. (2015). Iron nutrition, biomass production, and plant product quality. *Trends Plant Sci.* 20, 33–40. doi: 10.1016/j.tplants.2014.07.005
- Brown, J. C., and Chaney, R. L. (1971). Effect of iron on the transport of citrate into the xylem of soybeans and tomatoes. *Plant Physiol.* 47, 836–840. doi: 10.1104/pp.47.6.836
- Brüggemann, W., Maas-Kantel, K., and Moog, P. R. (1993). Iron uptake by leaf mesophyll cells: the role of the plasma membrane-bound ferric-chelate reductase. *Planta* 190, 151–155. doi: 10.1007/BF00196606
- Brumbarova, T., Bauer, P., and Ivanov, R. (2015). Molecular mechanisms governing *Arabidopsis* iron uptake. *Trends Plant Sci.* 20, 124–133. doi: 10.1016/j.tplants.2014.11.004
- Burton, J. W., Harlow, C., and Theil, E. C. (1998). Evidence for reutilization of nodule iron in soybean seed development. *J. Plant Nutr.* 21, 913–927. doi: 10.1080/01904169809365453
- Cakmak, I., Pfeiffer, W. H., and McClafferty, B. (2010). REVIEW: biofortification of durum wheat with zinc and iron. *Cereal Chem. J.* 87, 10–20. doi: 10.1094/CCHEM-87-1-0010
- Campion, B., Sparvoli, F., Doria, E., Tagliabue, G., Galasso, I., Fileppi, M., et al. (2009). Isolation and characterisation of an lpa (low phytic acid) mutant in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 118, 1211–1221. doi: 10.1007/s00122-009-0975-8
- Caracuta, V., Barzilay, O., Khalaili, H., Milevski, I., Paz, Y., Vardi, J., et al. (2015). The onset of faba bean farming in the Southern Levant. *Sci. Rep.* 5:14370. doi: 10.1038/srep14370
- Carpenter, C. E., and Mahoney, A. W. (1992). Contributions of heme and nonheme iron to human nutrition. *Crit. Rev. Food Sci. Nutr.* 31, 333–367. doi: 10.1080/10408399209527576
- Carvalho, L. C., Dennis, P. G., Fedoseyenko, D., Hajirezaei, M. R., Borriss, R., and von Wirén, N. (2011). Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency. *J. Plant Nutr. Soil Sci.* 174, 3–11. doi: 10.1002/jpln.201000085
- Cassin, G., Mari, S., Curie, C., Briat, J.-F., and Czernic, P. (2009). Increased sensitivity to iron deficiency in *Arabidopsis thaliana* overaccumulating nicotianamine. *J. Exp. Bot.* 60, 1249–1259. doi: 10.1093/jxb/erp007
- Cercamondi, C. I., Egli, I. M., Mitchikpe, E., Tossou, F., Zeder, C., Hounhouigan, J. D., et al. (2013). Total iron absorption by young women from iron-biofortified pearl millet composite meals is double that from regular millet meals but less than that from post-harvest iron-fortified millet meals. *J. Nutr.* 143, 1376–1382. doi: 10.3945/jn.113.176826
- Chitra, U., Singh, U., and Rao, P. V. (1997). Effect of varieties and processing methods on the total and ionizable iron contents of grain legumes. *J. Agric. Food Chem.* 45, 3859–3862. doi: 10.1021/jf970073a

- Chitra, U., Vimala, V., Singh, U., and Geervani, P. (1995). Variability in phytic acid content and protein digestibility of grain legumes. *Plant Foods Hum. Nutr.* 47, 163–172. doi: 10.1007/BF01089266
- Connolly, E. L., Campbell, N. H., Grotz, N., Prichard, C. L., and Guerinot, M. L. (2003). Overexpression of the FRO2 ferric chelate reductase confers tolerance to growth on low iron and uncovers posttranscriptional control. *Plant Physiol.* 133, 1102–1110. doi: 10.1104/pp.103.025122
- Connolly, E. L., Fett, J. P., and Guerinot, M. L. (2002). Expression of the IRT1 metal transporter is controlled by metals at the levels of transcript and protein accumulation. *Plant Cell* 14, 1347–1357. doi: 10.1105/tpc.001263
- Connorton, J. M., Balk, J., and Rodríguez-Celma, J. (2017). Iron homeostasis in plants – a brief overview. *Metallomics* 9, 813–823. doi: 10.1039/C7MT00136C
- Curie, C., and Mari, S. (2017). New routes for plant iron mining. *New Phytol.* 214, 521–525. doi: 10.1111/nph.14364
- Curie, C., Panaviene, Z., Loulergue, C., Dellaporta, S. L., Briat, J. F., and Walker, E. L. (2001). Maize *yellow stripe1* encodes a membrane protein directly involved in Fe(III) uptake. *Nature* 409, 346–349. doi: 10.1038/35053080
- Davila-Hicks, P., Theil, E. C., and Lönnerdal, B. (2004). Iron in ferritin or in salts (ferrous sulfate) is equally bioavailable in nonanemic women. *Am. J. Clin. Nutr.* 80, 936–940.
- De Moura, F. F., Palmer, A. C., Finkelstein, J. L., Haas, J. D., Murray-Kolb, L. E., Wenger, M. J., et al. (2014). Are biofortified staple food crops improving Vitamin A and iron status in women and children? New evidence from efficacy trials. *Adv. Nutr.* 5, 568–570. doi: 10.3945/an.114.006627
- Dell'Orto, M., Santi, S., De Nisi, P., Cesco, S., Varanini, Z., Zocchi, G., et al. (2000). Development of Fe-deficiency responses in cucumber (*Cucumis sativus* L.) roots: Involvement of plasma membrane H⁺-ATPase activity. *J. Exp. Bot.* 51, 695–701. doi: 10.1093/jexbot/51.345.695
- Diapari, M., Sindhu, A., Bett, K., Deokar, A., Warkentin, T. D., and Tar'an, B. (2014). Genetic diversity and association mapping of iron and zinc concentrations in chickpea (*Cicer arietinum* L.). *Genome* 57, 459–468. doi: 10.1139/gen-2014-0108
- Douchkov, D., Gryczka, C., Stephan, U. W., Hell, R., and Bäumlein, H. (2005). Ectopic expression of nicotianamine synthase genes results in improved iron accumulation and increased nickel tolerance in transgenic tobacco. *Plant Cell Environ.* 28, 365–374. doi: 10.1111/j.1365-3040.2005.01273.x
- Drakakaki, G., Marcel, S., Glahn, R., Lund, E., Pariagh, S., Fischer, R., et al. (2005). Endosperm-specific co-expression of recombinant soybean ferritin and *Aspergillus* phytase in maize results in significant increases in the levels of bioavailable iron. *Plant Mol. Biol.* 59, 869–880. doi: 10.1007/s11103-005-1537-3
- Eide, D., Broderius, M., Fett, J., and Guerinot, M. L. (1996). A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5624–5628. doi: 10.1073/pnas.93.11.5624
- FAO (1994). *Pulses and Derived Products*.
- FAO (2012). *The State of Food Insecurity in the World*. FAO Economic and Social Development Department, Rome.
- FAO (2014). *Food Outlook: Biannual Report on Global Food Markets*. FAO.
- FAO (2016a). *Annual Growth Rates of Total Pulse Production from 1961 to 2013*. Available online at: <http://faostat3.fao.org/browse/Q/QC/E>
- FAO (2016b). *FAO Statistical Database* (FAOSTAT).
- Finkelstein, J. L., Haas, J. D., and Mehta, S. (2017). Iron-biofortified staple food crops for improving iron status: a review of the current evidence. *Curr. Opin. Biotechnol.* 44, 138–145. doi: 10.1016/j.copbio.2017.01.003
- Finkelstein, J. L., Mehta, S., Udipi, S. A., Ghugre, P. S., Luna, S. V., Wenger, M. J., et al. (2015). A randomized trial of iron-biofortified pearl millet in school children in India. *J. Nutr.* 145, 1576–1581. doi: 10.3945/jn.114.208009
- Fuller, D. Q., Korisettar, R., and Venkatasubbiah, P. C. (2001). Southern Neolithic cultivation systems: a reconstruction based on archaeobotanical evidence. *South Asian Stud.* 17, 171–187. doi: 10.1080/02666030.2001.9628599
- Garnett, T. P., and Graham, R. D. (2005). Distribution and remobilization of iron and copper in wheat. *Ann. Bot.* 95, 817–826. doi: 10.1093/aob/mci085
- Gaur, P. M., Samineni, S., Sajja, S., and Chibbar, R. N. (2015). Achievements and challenges in improving nutritional quality of chickpea. *Legume Perspect.* 31–33.
- Gera, T., Sachdev, H. S., and Boy, E. (2012). Effect of iron-fortified foods on hematologic and biological outcomes: systematic review of randomized controlled trials. *Am. J. Clin. Nutr.* 96, 309–324. doi: 10.3945/ajcn.111.031500
- Ghavidel, R. A., and Prakash, J. (2007). The impact of germination and dehulling on nutrients, antinutrients, *in vitro* iron and calcium bioavailability and *in vitro* starch and protein digestibility of some legume seeds. *LWT Food Sci. Technol.* 40, 1292–1299. doi: 10.1016/j.lwt.2006.08.002
- Gooding, M. J., Kasyanova, E., Ruske, R., Hauggaard-Nielsen, H., Jensen, E. S., Dahlmann, C., et al. (2007). Intercropping with pulses to concentrate nitrogen and sulphur in wheat. *J. Agric. Sci.* 145, 469–479. doi: 10.1017/S0021859607007241
- Goto, F., Yoshihara, T., and Saiki, H. (2000). Iron accumulation and enhanced growth in transgenic lettuce plants expressing the iron-binding protein ferritin. *Theor. Appl. Genet.* 100, 658–664. doi: 10.1007/s001220051336
- Griffiths, D. W., and Thomas, T. A. (1981). Phytate and total phosphorus content of field beans (*Vicia faba* L.). *J. Sci. Food Agric.* 32, 187–192. doi: 10.1002/jsfa.2740320215
- Grillet, L., Ouerdane, L., Flis, P., Hoang, M. T. T., Isaure, M.-P., Lobinski, R., et al. (2014). Ascorbate efflux as a new strategy for iron reduction and transport in plants. *J. Biol. Chem.* 289, 2515–2525. doi: 10.1074/jbc.M113.514828
- Guttieri, M., Bowen, D., Dorsch, J. A., Raboy, V., and Souza, E. (2004). Identification and characterization of a low phytic acid wheat. *Crop Sci.* 44, 418–424. doi: 10.2135/cropsci2004.4180
- Haas, J. D., Beard, J. L., Murray-Kolb, L. E., del Mundo, A. M., Felix, A., and Gregorio, G. B. (2005). Iron-biofortified rice improves the iron stores of nonanemic Filipino women. *J. Nutr.* 135, 2823–2830.
- Haber, F., and Weiss, J. (1934). The catalytic decomposition of hydrogen peroxide by iron salts. *Proc. R. Soc. Lond. A* 147, 332–351. doi: 10.1098/rspa.1934.0221
- Hallberg, L., Brune, M., and Rossander, L. (1989). Iron absorption in man: ascorbic acid and dose-dependent inhibition by phytate. *Am. J. Clin. Nutr.* 49, 140–144.
- HarvestPlus (2015a). *Crops*.
- HarvestPlus (2015b). *FAQ about Biofortification*.
- Hell, R., and Stephan, U. W. (2003). Iron uptake, trafficking and homeostasis in plants. *Planta* 216, 541–551. doi: 10.1007/s00425-002-0920-4
- Hemalatha, S., Platel, K., and Srinivasan, K. (2007a). Influence of heat processing on the bioaccessibility of zinc and iron from cereals and pulses consumed in India. *J. Trace Elem. Med. Biol.* 21, 1–7. doi: 10.1016/j.jtemb.2006.10.002
- Hemalatha, S., Platel, K., and Srinivasan, K. (2007b). Zinc and iron contents and their bioaccessibility in cereals and pulses consumed in India. *Food Chem.* 102, 1328–1336. doi: 10.1016/j.foodchem.2006.07.015
- Higuchi, K., Kanazawa, K., Nishizawa, N.-K., Chino, M., and Mori, S. (1994). Purification and characterization of nicotianamine synthase from Fe-deficient barley roots. *Plant Soil* 165, 173–179. doi: 10.1007/BF00008059
- Higuchi, K., Nakanishi, H., Suzuki, K., Nishizawa, N. K., and Mori, S. (1999). Presence of nicotianamine synthase isozymes and their homologues in the root of graminaceous plants. *Soil Sci. Plant Nutr.* 45, 681–691. doi: 10.1080/00380768.1999.10415831
- Higuchi, K., Nishizawa, N.-K., Yamaguchi, H., Römhelt, V., Marschner, H., and Mori, S. (1995). SHORT COMMUNICATION: Response of nicotianamine synthase activity to Fe-deficiency in tobacco plants as compared with barley. *J. Exp. Bot.* 46, 1061–1063. doi: 10.1093/jxb/46.8.1061
- Hocking, P. J., and Pate, J. S. (1977). Mobilization of minerals to developing seeds of legumes. *Ann. Bot.* 41, 1259–1278. doi: 10.1093/oxfordjournals.aob.a085415
- Horton, S., Mannar, V., and Wesley, A. (2008). “Micronutrient fortification (iron and salt iodization),” in *Copenhagen Consensus 2008* (Denmark).
- House, W. A. (1999). Trace element bioavailability as exemplified by iron and zinc. *Field Crops Res.* 60, 115–141. doi: 10.1016/S0378-4290(98)00136-1
- Hurrell, R. F., Reddy, M. B., Juillerat, M.-A., and Cook, J. D. (2003). Degradation of phytic acid in cereal porridges improves iron absorption by human subjects. *Am. J. Clin. Nutr.* 77, 1213–1219.
- Ibrikci, H., Knewton, S. J. B., and Grusak, M. A. (2003). Chickpea leaves as a vegetable green for humans: evaluation of mineral composition. *J. Sci. Food Agric.* 83, 945–950. doi: 10.1002/jsfa.1427
- Ihemere, U. E., Narayanan, N. N., and Sayre, R. T. (2012). Iron biofortification and homeostasis in transgenic cassava roots expressing the algal iron assimilatory gene, *FEA1*. *Front. Plant Sci.* 3:71. doi: 10.3389/fpls.2012.00171
- IMECHE (2013). *Global Food - Waste Not, Want Not*. Wales.
- Indurker, S., Misra, H. S., and Eapen, S. (2010). *Agrobacterium*-mediated transformation in chickpea (*Cicer arietinum* L.) with an insecticidal protein gene: optimisation of different factors. *Physiol. Mol. Biol. Plants* 16, 273–284. doi: 10.1007/s12298-010-0030-x

- Inoue, H., Kobayashi, T., Nozoye, T., Takahashi, M., Kakei, Y., Suzuki, K., et al. (2009). Rice OsYSL15 is an iron-regulated iron (III)-deoxymugineic acid transporter expressed in the roots and is essential for iron uptake in early growth of the seedlings. *J. Biol. Chem.* 284, 3470–3479. doi: 10.1074/jbc.M806042200
- Iqbal, A., Khalil, I. A., Ateeq, N., and Sayyar Khan, M. (2006). Nutritional quality of important food legumes. *Food Chem.* 97, 331–335. doi: 10.1016/j.foodchem.2005.05.011
- Ishimaru, Y., Masuda, H., Bashir, K., Inoue, H., Tsukamoto, T., Takahashi, M., et al. (2010). Rice metal-nicotianamine transporter, OsYSL2, is required for the long-distance transport of iron and manganese. *Plant J.* 62, 379–390. doi: 10.1111/j.1365-313X.2010.04158.x
- Ishimaru, Y., Suzuki, M., Tsukamoto, T., Suzuki, K., Nakazono, M., Kobayashi, T., et al. (2006). Rice plants take up iron as Fe^{3+} -phytosiderophore and as Fe^{2+} . *Plant J.* 45, 335–346. doi: 10.1111/j.1365-313X.2005.02624.x
- Jeong, J., and Connolly, E. L. (2009). Iron uptake mechanisms in plants: functions of the FRO family of ferric reductases. *Plant Sci.* 176, 709–714. doi: 10.1016/j.plantsci.2009.02.011
- Jeong, J., and Gueriot, M. L. (2009). Homing in on iron homeostasis in plants. *Trends Plant Sci.* 14, 280–285. doi: 10.1016/j.tplants.2009.02.006
- Jeong, J., Cohu, C., Kerkeb, L., Pilon, M., Connolly, E. L., and Gueriot, M. L. (2008). Chloroplast Fe(III) chelate reductase activity is essential for seedling viability under iron limiting conditions. *Proc. Natl. Acad. Sci. U.S.A.* 105, 10619–10624. doi: 10.1073/pnas.0708367105
- Jin, T., Zhou, J., Chen, J., Zhu, L., Zhao, Y., and Huang, Y. (2013). The genetic architecture of zinc and iron content in maize grains as revealed by QTL mapping and meta-analysis. *Breed. Sci.* 63, 317–324. doi: 10.1270/jsbbs.63.317
- Johnson, A. A. T., Kyriacou, B., Callahan, D. L., Carruthers, L., Stangoulis, J., Lombi, E., et al. (2011). Constitutive overexpression of the OsNAS gene family reveals single-gene strategies for effective iron- and zinc-biofortification of rice endosperm. *PLoS ONE* 6:e24476. doi: 10.1371/journal.pone.0024476
- Jones, D. L., Darah, P. R., and Kochian, L. V. (1996). Critical evaluation of organic acid mediated iron dissolution in the rhizosphere and its potential role in root iron uptake. *Plant Soil* 180, 57–66. doi: 10.1007/BF00015411
- Jukanti, A. K., Gaur, P. M., Gowda, C. L. L., and Chibbar, R. N. (2012). Nutritional quality and health benefits of chickpea (*Cicer arietinum* L.): a review. *Br. J. Nutr.* 108, S11–S26. doi: 10.1017/S0007114512000797
- Kanazawa, K., Higuchi, K., Nishizawa, N.-K., Fushiya, S., Chino, M., and Mori, S. (1994). Nicotianamine aminotransferase activities are correlated to the phytosiderophore secretions under Fe-deficient conditions in Gramineae. *J. Exp. Bot.* 45, 1903–1906. doi: 10.1093/jxb/45.12.1903
- Kassebaum, N. J. (2016). The global burden of anemia. *Hematol. Oncol. Clin. North Am.* 30, 247–308. doi: 10.1016/j.hoc.2015.11.002
- Kataki, P. K. (2002). Shifts in cropping system and its effect on human nutrition: case study from India. *J. Crop Product.* 6, 119–144. doi: 10.1300/J144v06n01_08
- Katsvairo, L. (2015). “Delivery of iron beans in Rwanda,” in *The 2nd Global Conference on Biofortification: Getting Nutritious Foods to People* (Kigali).
- Kawai, S., Kamei, S., Matsuda, Y., Ando, R., Kondo, S., Ishizawa, A., et al. (2001). Concentrations of iron and phytosiderophores in xylem sap of iron-deficient barley plants. *Soil Sci. Plant Nutr.* 47, 265–272. doi: 10.1080/00380768.2001.10408390
- Kehrer, J. P. (2000). The Haber–Weiss reaction and mechanisms of toxicity. *Toxicology* 149, 43–50. doi: 10.1016/S0300-483X(00)00231-6
- Kennedy, E. T., and Bouis, H. E. (1993). *Linkages between Agriculture and Nutrition: Implications for Policy and Research*. International Food Policy Research Institute, Washington, USA.
- Kim, S. A., and Gueriot, M. L. (2007). Mining iron: iron uptake and transport in plants. *FEBS Lett.* 581, 2273–2280. doi: 10.1016/j.febslet.2007.04.043
- Kobayashi, T., and Nishizawa, N. K. (2012). Iron uptake, translocation, and regulation in higher plants. *Annu. Rev. Plant Biol.* 63, 131–152. doi: 10.1146/annurev-arplant-042811-105522
- Kodany, B. S., Bellad, R. M., Mahantshetti, N. S., Westcott, J. E., Krebs, N. F., Kemp, J. F., et al. (2013). Biofortification of pearl millet with iron and zinc in a randomized controlled trial increases absorption of these minerals above physiologic requirements in young children. *J. Nutr.* 143, 1489–1493. doi: 10.3945/jn.113.176677
- Koike, S., Inoue, H., Mizuno, D., Takahashi, M., Nakanishi, H., Mori, S., et al. (2004). OsYSL2 is a rice metal-nicotianamine transporter that is regulated by iron and expressed in the phloem. *Plant J.* 39, 415–424. doi: 10.1111/j.1365-313X.2004.02146.x
- Kon, S., and Sanchuck, D. W. (1981). Phytate content and its effect on cooking quality of beans. *J. Food Process. Preserv.* 5, 169–178. doi: 10.1111/j.1745-4549.1981.tb00632.x
- Kumar, S., Hash, C. T., Thirunavukkarasu, N., Singh, G., Rajaram, V., Rathore, A., et al. (2016). Mapping quantitative trait loci controlling high iron and zinc content in self and open pollinated grains of pearl millet [*Pennisetum glaucum* (L.) R. Br.]. *Front. Plant Sci.* 7:1636. doi: 10.3389/fpls.2016.01636
- Lanquar, V., Lelièvre, F., Bolte, S., Hamès, C., Alcon, C., Neumann, D., et al. (2005). Mobilization of vacuolar iron by AtNRAMP3 and AtNRAMP4 is essential for seed germination on low iron. *EMBO J.* 24, 4041–4051. doi: 10.1038/sj.emboj.7600864
- Larson, S. R., Rutger, J. N., Young, K. A., and Raboy, V. (2000). Isolation and genetic mapping of a non-lethal rice (*Oryza sativa* L.) low phytic acid 1 mutation. *Crop Sci.* 40, 1397–1405. doi: 10.2135/cropsci2000.4051397x
- Larson, S. R., Young, K. A., Cook, A., Blake, T. K., and Raboy, V. (1998). Linkage mapping of two mutations that reduce phytic acid content of barley grain. *Theor. Appl. Genet.* 97, 141–146. doi: 10.1007/s001220050878
- Lee, S., Kim, Y.-S., Jeon, U., Kim, Y.-K., Schjoerring, J., and An, G. (2012). Activation of rice nicotianamine synthase 2 (OsNAS2) enhances iron availability for biofortification. *Mol. Cells* 33, 269–275. doi: 10.1007/s10059-012-2231-3
- Lee, S., Persson, D. P., Hansen, T. H., Husted, S., Schjoerring, J. K., Kim, Y.-S., et al. (2011). Bio-available zinc in rice seeds is increased by activation tagging of nicotianamine synthase. *Plant Biotechnol. J.* 9, 865–873. doi: 10.1111/j.1467-7652.2011.00606.x
- Li, L., Cheng, X., and Ling, H.-Q. (2004a). Isolation and characterization of Fe(III)-chelate reductase gene *LeFRO1* in tomato. *Plant Mol. Biol.* 54, 125–136. doi: 10.1023/B:PLAN.0000028774.82782.16
- Li, L., Tang, C., Rengel, Z., and Zhang, F. (2003). Chickpea facilitates phosphorus uptake by intercropped wheat from an organic phosphorus source. *Plant Soil* 248, 297–303. doi: 10.1023/A:1022389707051
- Li, S. M., Li, L., Zhang, F. S., and Tang, C. (2004b). Acid phosphatase role in chickpea/maize intercropping. *Ann. Bot.* 94, 297–303. doi: 10.1093/aob/mch140
- Li, S., Zhou, X., Li, H., Liu, Y., Zhu, L., Guo, J., et al. (2015). Overexpression of *ZmIRT1* and *ZmZIP3* enhances iron and zinc accumulation in transgenic *Arabidopsis*. *PLoS ONE* 10: e0136647. doi: 10.1371/journal.pone.0136647
- Lin, Y.-F., Liang, H.-M., Yang, S.-Y., Boch, A., Clemens, S., Chen, C.-C., et al. (2009). Arabidopsis IRT3 is a zinc-regulated and plasma membrane localized zinc/iron transporter. *New Phytol.* 182, 392–404. doi: 10.1111/j.1469-8137.2009.02766.x
- Lolas, G., Palamidis, N., and Markakis, P. (1976). The phytic acid-total phosphorus relationship in barley, oats, soybeans, and wheat. *Cereal Chem. J.* 53, 867–871.
- Lönnnerdal, B., Bryant, A., Liu, X., and Theil, E. C. (2006). Iron absorption from soybean ferritin in nonanemic women. *Am. J. Clin. Nutr.* 83, 103–107.
- Lucca, P., Hurrell, R., and Potrykus, I. (2001). Genetic engineering approaches to improve the bioavailability and the level of iron in rice grains. *Theor. Appl. Genet.* 102, 392–397. doi: 10.1007/s001220051659
- Ma, J. F., and Nomoto, K. (1993). Two related biosynthetic pathways of mugineic acids in gramineous plants. *Plant Physiol.* 102, 373–378. doi: 10.1104/pp.102.2.373
- Mahmoudi, H., Labidi, N., Ksouri, R., Gharsalli, M., and Abdelly, C. (2007). Differential tolerance to iron deficiency of chickpea varieties and Fe resupply effects. *C. R. Biol.* 330, 237–246. doi: 10.1016/j.crv.2007.02.007
- Maillard, A., Diquélou, S., Billard, V., Lainé, P., Garnica, M., Prudent, M., et al. (2015). Leaf mineral nutrient remobilization during leaf senescence and modulation by nutrient deficiency. *Front. Plant Sci.* 6:317. doi: 10.3389/fpls.2015.00317
- Masuda, H., Aung, M., and Nishizawa, N. (2013a). Iron biofortification of rice using different transgenic approaches. *Rice* 6, 1–12. doi: 10.1186/1939-8433-6-40
- Masuda, H., Ishimaru, Y., Aung, M. S., Kobayashi, T., Kakei, Y., Takahashi, M., et al. (2012). Iron biofortification in rice by the introduction of multiple genes involved in iron nutrition. *Sci. Rep.* 2:543. doi: 10.1038/srep00543
- Masuda, H., Kobayashi, T., Ishimaru, Y., Takahashi, M., Aung, M. S., Nakanishi, H., et al. (2013b). Iron-biofortification in rice by the introduction of three barley

- genes participated in mugineic acid biosynthesis with soybean *ferritin* gene. *Front. Plant Sci.* 4:132. doi: 10.3389/fpls.2013.00132
- Masuda, H., Usuda, K., Kobayashi, T., Ishimaru, Y., Kakei, Y., Takahashi, M., et al. (2009). Overexpression of the barley nicotianamine synthase gene *HvNAS1* increases iron and zinc concentrations in rice grains. *Rice* 2, 155–166. doi: 10.1007/s12284-009-9031-1
- Matovu, M. (2016). *Molecular and Biochemical Characterisation of Transgenic Banana Lines Containing Iron Uptake and Storage Enhancing Genes*. Ph.D. Queensland University of Technology.
- Meiners, C. R., Derise, N. L., Lau, H. C., Crews, M. G., Ritchey, S. J., and Murphy, E. W. (1976). The content of nine mineral elements in raw and cooked mature dry legumes. *J. Agric. Food Chem.* 24, 1126–1130. doi: 10.1021/jf60208a036
- Mendoza, C., Viteri, F. E., Lönnnerdal, B., Young, K. A., Raboy, V., and Brown, K. H. (1998). Effect of genetically modified, low-phytic acid maize on absorption of iron from tortillas. *Am. J. Clin. Nutr.* 68, 1123–1127.
- Mishra, P. K., Bisht, S. C., Ruwari, P., Joshi, G. K., Singh, G., Bisht, J. K., et al. (2011). Bioassociative effect of cold tolerant *Pseudomonas* spp. and *Rhizobium leguminosarum*-PR1 on iron acquisition, nutrient uptake and growth of lentil (*Lens culinaris* L.). *Eur. J. Soil Biol.* 47, 35–43. doi: 10.1016/j.ejsobi.2010.11.005
- Mori, S., and Nishizawa, N. (1989). Identification of barley chromosome no. 4, possible encoder of genes of mugineic acid synthesis from 2'-deoxymugineic acid using wheat-barley addition lines. *Plant Cell Physiol.* 30, 1057–1061.
- Moriyama, M., and Oba, K. (2008). Comparative study on the vitamin C contents of the food legume seeds. *J. Nutr. Sci. Vitaminol.* 54, 1–6. doi: 10.3177/jnsv.54.1
- Morrissey, J., and Guerinot, M. L. (2009). Iron uptake and transport in plants: the good, the bad, and the ionome. *Chem. Rev.* 109, 4553–4567. doi: 10.1021/cr900112r
- Munyaka, A. W., Makule, E. E., Oey, I., Van Loey, A., and Hendrickx, M. (2010). Thermal stability of L-ascorbic acid and ascorbic acid oxidase in broccoli (*Brassica oleracea* var. *italica*). *J. Food Sci.* 75, C336–C340. doi: 10.1111/j.1750-3841.2010.01573.x
- Murata, Y., Ma, J. F., Yamaji, N., Ueno, D., Nomoto, K., and Iwashita, T. (2006). A specific transporter for iron(III)-phytosiderophore in barley roots. *Plant J.* 46, 563–572. doi: 10.1111/j.1365-3113X.2006.02714.x
- Namanya, P. (2011). *Towards the Biofortification of Banana Fruit for Enhanced Micronutrient Content*. Ph.D. Queensland University of Technology.
- Narayanan, N., Beyene, G., Chauhan, R. D., Gaitán-Solis, E., Grusak, M. A., Taylor, N., et al. (2015). Overexpression of *Arabidopsis* VIT1 increases accumulation of iron in cassava roots and stems. *Plant Sci.* 240, 170–181. doi: 10.1016/j.plantsci.2015.09.007
- Nelson, R. L., Yoo, S. J., Tanure, J. C., Andrianopoulos, G., and Misumi, A. (1988). The effect of iron on experimental colorectal carcinogenesis. *Anticancer Res.* 9, 1477–1482.
- Nestel, P., Bouis, H. E., Meenakshi, J. V., and Pfeiffer, W. (2006). Biofortification of staple food crops. *J. Nutr.* 136, 1064–1067.
- Ning, L., Sun, P., Wang, Q., Ma, D., Hu, Z., Zhang, D., et al. (2015). Genetic architecture of biofortification traits in soybean (*Glycine max* L. Merr.) revealed through association analysis and linkage mapping. *Euphytica* 204, 353–369. doi: 10.1007/s10681-014-1340-9
- Norton, G. J., Deacon, C. M., Xiong, L., Huang, S., Meharg, A. A., and Price, A. H. (2010). Genetic mapping of the rice ionome in leaves and grain: identification of QTLs for 17 elements including arsenic, cadmium, iron and selenium. *Plant Soil* 329, 139–153. doi: 10.1007/s11104-009-0141-8
- Nozoye, T., Nagasaka, S., Kobayashi, T., Takahashi, M., Sato, Y., Sato, Y., et al. (2011). Phytosiderophore efflux transporters are crucial for iron acquisition in graminaceous plants. *J. Biol. Chem.* 286, 5446–5454. doi: 10.1074/jbc.M110.180026
- Pahlavan-Rad, M. R., and Pessarakli, M. (2009). Response of wheat plants to zinc, iron, and manganese applications and uptake and concentration of zinc, iron, and manganese in wheat grains. *Commun. Soil Sci. Plant Anal.* 40, 1322–1332. doi: 10.1080/00103620902761262
- Papanikolaou, G., and Pantopoulos, K. (2005). Iron metabolism and toxicity. *Toxicol. Appl. Pharmacol.* 202, 199–211. doi: 10.1016/j.taap.2004.06.021
- Peng, A. H., Liu, X. F., He, Y. R., Xu, L. Z., Lei, T. G., Yao, L. X., et al. (2015). Characterization of transgenic *Poncirus trifoliata* overexpressing the ferric chelate reductase gene *CjFRO2* from *Citrus junos*. *Biol. Plant.* 59, 654–660. doi: 10.1007/s10535-015-0543-9
- Petry, N., Egli, I., Campion, B., Nielsen, E., and Hurrell, R. (2013). Genetic reduction of phytate in common bean (*Phaseolus vulgaris* L.) seeds increases iron absorption in young women. *J. Nutr.* 143, 1219–1224. doi: 10.3945/jn.113.175067
- Petry, N., Egli, I., Gahutu, J. B., Tugirimana, P. L., Boy, E., and Hurrell, R. (2012). Stable iron isotope studies in Rwandese women indicate that the common bean has limited potential as a vehicle for iron biofortification. *J. Nutr.* 142, 492–497. doi: 10.3945/jn.111.149286
- Petry, N., Egli, I., Gahutu, J. B., Tugirimana, P. L., Boy, E., and Hurrell, R. (2014). Phytic acid concentration influences iron bioavailability from biofortified beans in Rwandese women with low iron status. *J. Nutr.* 144, 1681–1687. doi: 10.3945/jn.114.192989
- Petry, N., Rohner, F., Gahutu, J. B., Campion, B., Boy, E., Tugirimana, P. L., et al. (2016). In Rwandese women with low iron status, iron absorption from low-phytic acid beans and biofortified beans is comparable, but low-phytic acid beans cause adverse gastrointestinal symptoms. *J. Nutr.* 146, 970–975. doi: 10.3945/jn.115.223693
- Phillips, R. D. (1993). Starchy legumes in human nutrition, health and culture. *Plant Foods Hum. Nutr.* 44, 195–211. doi: 10.1007/BF01088314
- Pianelli, K., Mari, S., Marquès, L., Lebrun, M., and Czernic, P. (2005). Nicotianamine over-accumulation confers resistance to nickel in *Arabidopsis thaliana*. *Transgenic Res.* 14, 739–748. doi: 10.1007/s11248-005-7159-3
- Pich, A., Scholz, G., and Stephan, U. (1994). Iron-dependent changes of heavy metals, nicotianamine, and citrate in different plant organs and in the xylem exudate of two tomato genotypes. Nicotianamine as possible copper translocator. *Plant Soil* 165, 189–196. doi: 10.1007/BF00008061
- Pietrangelo, A. (2003). Haemochromatosis. *Gut* 52, ii23–ii30. doi: 10.1136/gut.52.suppl_2.ii23
- Pingali, P. L. (2012). Green revolution: impacts, limits, and the path ahead. *Proc. Natl. Acad. Sci. U.S.A.* 109, 12302–12308. doi: 10.1073/pnas.0912953109
- Pinstrip-Andersen, P., and Hazell, P. B. (1985). The impact of the Green Revolution and prospects for the future. *Food Rev. Int.* 1, 1–25. doi: 10.1080/87559128509540765
- Platel, K., and Srinivasan, K. (2016). Bioavailability of micronutrients from plant foods: an update. *Crit. Rev. Food Sci. Nutr.* 56, 1608–1619. doi: 10.1080/10408398.2013.781011
- Poltronieri, F., Aréas, J. A. G., and Colli, C. (2000). Extrusion and iron bioavailability in chickpea (*Cicer arietinum* L.). *Food Chem.* 70, 175–180. doi: 10.1016/S0956-7135(99)00113-9
- Pulse Australia (2016a) Best Management Guide. Chickpea Production: *Northern Region*.
- Pulse Australia (2016b) *What are Pulses?*
- Pulse Canada (2016). *Protein Quality of Cooked Pulses (PDCAAS Method)*.
- Rabotti, G., and Zocchi, G. (1994). Plasma membrane-bound H⁺-ATPase and reductase activities in Fe-deficient cucumber roots. *Physiol. Plant.* 90, 779–785. doi: 10.1111/j.1399-3054.1994.tb02537.x
- Raboy, V., Gerbasi, P. F., Young, K. A., Stoneberg, S. D., Pickett, S. G., Bauman, A. T., et al. (2000). Origin and seed phenotype of maize low phytic acid 1-1 and low phytic acid 2-1. *Plant Physiol.* 124, 355–368. doi: 10.1104/pp.124.1.355
- Rana, A., Saharan, B., Nain, L., Prasanna, R., and Shivay, Y. S. (2012). Enhancing micronutrient uptake and yield of wheat through bacterial PGPR consortia. *Soil Sci. Plant Nutr.* 58, 573–582. doi: 10.1080/00380768.2012.716750
- Rao, M., and Willey, R. (1980). Evaluation of yield stability in intercropping: studies on sorghum/pigeonpea. *Exp. Agric.* 16, 105–116. doi: 10.1017/S0014479700010796
- Rao, P. P., Bithal, P. S., Bhagavatula, S., and Bantilan, M. C. S. (2010). *Chickpea and Pigeonpea Economies in Asia: Facts, Trends and Outlook*. Patancheru: International Crops Research Institute for the Semi-Arid Tropics.
- Rao, S., Joshi, S., Bhide, P., Puranik, B., and Asawari, K. (2013). Dietary diversification for prevention of anaemia among women of childbearing age from rural India. *Public Health Nutr.* 17, 939–947. doi: 10.1017/S1368980013001006
- Ravindran, V., Ravindran, G., and Sivalogan, S. (1994). Total and phytate phosphorus contents of various foods and feedstuffs of plant origin. *Food Chem.* 50, 133–136. doi: 10.1016/0308-8146(94)90109-0
- Ray, H., Bett, K., Tar'an, B., Vandenberg, A., Thavarajah, D., and Warkentin, T. (2014). Mineral micronutrient content of cultivars of field pea, chickpea,

- common bean, and lentil grown in Saskatchewan, Canada. *Crop Sci.* 54, 1698–1708. doi: 10.2135/cropsci2013.08.0568
- Rizkalla, S. W., Bellisle, F., and Slama, G. (2002). Health benefits of low glycaemic index foods, such as pulses, in diabetic patients and healthy individuals. *Br. J. Nutr.* 88, 255–262. doi: 10.1079/BJN2002715
- Robinson, N. J., Procter, C. M., Connolly, E. L., and Guerinot, M. L. (1999). A ferric-chelate reductase for iron uptake from soils. *Nature* 397, 694–697. doi: 10.1038/17800
- Rochfort, S., and Panozzo, J. (2007). Phytochemicals for health, the role of pulses. *J. Agric. Food Chem.* 55, 7981–7994. doi: 10.1021/jf071704w
- Rodríguez-Celma, J., Lin, W.-D., Fu, G.-M., Abadía, J., López-Millán, A.-F., and Schmidt, W. (2013). Mutually exclusive alterations in secondary metabolism are critical for the uptake of insoluble iron compounds by *Arabidopsis* and *Medicago truncatula*. *Plant Physiol.* 162, 1473–1485. doi: 10.1104/pp.113.220426
- Römhelt, V., and Marschner, H. (1981). Iron deficiency stress induced morphological and physiological changes in root tips of sunflower. *Physiol. Plant.* 53, 354–360. doi: 10.1111/j.1399-3054.1981.tb04512.x
- Römhelt, V., and Marschner, H. (1986). Evidence for a specific uptake system for iron phytosiderophores in roots of grasses. *Plant Physiol.* 80, 175–180. doi: 10.1104/pp.80.1.175
- Roy, F., Boye, J. I., and Simpson, B. K. (2010). Bioactive proteins and peptides in pulse crops: pea, chickpea and lentil. *Food Res. Int.* 43, 432–442. doi: 10.1016/j.foodres.2009.09.002
- Sánchez-Rodríguez, A. R., del Campillo, M. C., Torrent, J., and Jones, D. L. (2014). Organic acids alleviate iron chlorosis in chickpea grown on two P-fertilized soils. *J. Soil Sci. Plant Nutr.* 14, 292–303. doi: 10.4067/S0718-95162014005000024
- Sandberg, A.-S. (2002). Bioavailability of minerals in legumes. *Br. J. Nutr.* 88, 281–285. doi: 10.1079/BJN/2002718
- Sandberg, A.-S., Brune, M., Carlsson, N.-G., Hallberg, L., Skoglund, E., and Rossander-Hulthén, L. (1999). Inositol phosphates with different numbers of phosphate groups influence iron absorption in humans. *Am. J. Clin. Nutr.* 70, 240–246.
- Sandberg, A. S., Carlsson, N. G., and Svanberg, U. (1989). Effects of inositol tri-, tetra-, penta-, and hexaphosphates on *in vitro* estimation of iron availability. *J. Food Sci.* 54, 159–161. doi: 10.1111/j.1365-2621.1989.tb08591.x
- Sander, J. D., and Joung, J. K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotech.* 32, 347–355. doi: 10.1038/nbt.2842
- Santi, S., and Schmidt, W. (2009). Dissecting iron deficiency-induced proton extrusion in *Arabidopsis* roots. *New Phytol.* 183, 1072–1084. doi: 10.1111/j.1469-8137.2009.02908.x
- Santi, S., Cesco, S., Varanini, Z., and Pinton, R. (2005). Two plasma membrane H⁺-ATPase genes are differentially expressed in iron-deficient cucumber plants. *Plant Physiol. Biochem.* 43, 287–292. doi: 10.1016/j.plaphy.2005.02.007
- Santos, C. F., and Boiteux, L. (2015). Genetic control and transgressive segregation of zinc, iron, potassium, phosphorus, calcium, and sodium accumulation in cowpea (*Vigna unguiculata*) seeds. *Genetics Mol. Res.* 14, 259–268. doi: 10.4238/2015.January.16.10
- Sarmah, B., Moore, A., Tate, W., Molvig, L., Morton, R., Rees, D., et al. (2004). Transgenic chickpea seeds expressing high levels of a bean α -amylase inhibitor. *Mol. Breed.* 14, 73–82. doi: 10.1023/B:MOLB.0000037996.01494.12
- Sebastiá, V., Barberá, R., Farré, R., and Lagarda, M. J. (2001). Effects of legume processing on calcium, iron and zinc contents and dialysabilities. *J. Sci. Food Agric.* 81, 1180–1185. doi: 10.1002/jsfa.927
- Sharma, A., Shankhdhar, D., and Shankhdhar, S. (2013). Enhancing grain iron content of rice by the application of plant growth promoting rhizobacteria. *Plant Soil Environ.* 59, 89–94.
- Sharma, R. D. (1986). Phytate and the epidemiology of heart disease, renal calculi and colon cancer. *Phytic Acid Chem. Appl.* 161–172.
- Shi, R., Weber, G., Köster, J., Reza-Hajirezaei, M., Zou, C., Zhang, F., et al. (2012). Senescence-induced iron mobilization in source leaves of barley (*Hordeum vulgare*) plants. *New Phytol.* 195, 372–383. doi: 10.1111/j.1469-8137.2012.04165.x
- Siegenberg, D., Baynes, R. D., Bothwell, T. H., Macfarlane, B. J., Lamparelli, R. D., Car, N. G., et al. (1991). Ascorbic acid prevents the dose-dependent inhibitory effects of polyphenols and phytates on nonheme-iron absorption. *Am. J. Clin. Nutr.* 53, 537–541.
- Sievenpiper, J. L., Kendall, C. W. C., Esfahani, A., Wong, J. M. W., Carleton, A. J., Jiang, H. Y., et al. (2009). Effect of non-oil-seed pulses on glycaemic control: a systematic review and meta-analysis of randomised controlled experimental trials in people with and without diabetes. *Diabetologia* 52, 1479–1495. doi: 10.1007/s00125-009-1395-7
- Sood, M., and Malhotra, S. R. (2002). Effects of processing and cooking on ascorbic acid content of chickpea (*Cicer arietinum* L) varieties. *J. Sci. Food Agric.* 82, 65–68. doi: 10.1002/jsfa.1001
- Sperotto, R. A., Ricachenevsky, F. K., Waldow Vde, A., and Fett, J. P. (2012). Iron biofortification in rice: it's a long way to the top. *Plant Sci.* 190, 24–39. doi: 10.1016/j.plantsci.2012.03.004
- Stephan, U. W., and Scholz, G. (1993). Nicotianamine: mediator of transport of iron and heavy metals in the phloem? *Physiol. Plant.* 88, 522–529. doi: 10.1111/j.1399-3054.1993.tb01367.x
- Strozycki, P. M., Szczurek, A., Lotocka, B., Figlerowicz, M., and Legocki, A. B. (2007). Ferritins and nodulation in *Lupinus luteus*: iron management in indeterminate type nodules. *J. Exp. Bot.* 58, 3145–3153. doi: 10.1093/jxb/erm152
- Thavarajah, D., Thavarajah, P., and Gupta, D. S. (2014). “Pulses biofortification in genomic era: multidisciplinary opportunities and challenges,” in *Legumes in the Omic Era*, eds S. Gupta, N. Nadarajan, and S. D. Gupta (New York, NY: Springer New York), 207–220. doi: 10.1007/978-1-4614-8370-0_10
- Thomine, S., and Lanquar, V. (2011). Iron transport and signaling in plants. in *Transporters and Pumps in Plant Signaling*, eds M. Geisler, M., and K. Venema (Berlin; Heidelberg: Springer Berlin Heidelberg), 99–131. doi: 10.1007/978-3-642-14369-4_4
- Thomine, S., and Vert, G. (2013). Iron transport in plants: better be safe than sorry. *Curr. Opin. Plant Biol.* 16, 322–327. doi: 10.1016/j.pbi.2013.01.003
- Trijatmiko, K. R., Dueñas, C., Tsakirpaloglou, N., Torrizo, L., Arines, F. M., Adeva, C., et al. (2016). Biofortified indica rice attains iron and zinc nutrition dietary targets in the field. *Sci. Rep.* 6:19792. doi: 10.1038/srep19792
- Trumbo, P., Yates, A. A., Schlicker, S., and Poos, M. (2001). Dietary reference intakes: vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. *J. Am. Diet. Assoc.* 101, 294–301. doi: 10.1016/S0002-8223(01)00078-5
- Umaid, U., Manohar, S., and Singh, A. (1984). The anatomical structure of desi and kabuli chickpea seed coats. *Int. Chickpea Newsl.* 10, 26–27.
- United Nations (2015). *World Population Prospects, The 2015 Revision*. United Nations Department of Economic and Social Affairs.
- Upadhyaya, H. D., Bajaj, D., Das, S., Kumar, V., Gowda, C. L. L., Sharma, S., et al. (2016). Genetic dissection of seed-iron and zinc concentrations in chickpea. *Sci. Rep.* 6:24050. doi: 10.1038/srep24050
- USDA (2013). National Nutrient Database for Standard Reference.
- Van den Broeck, I., Ludikhuyze, L., Weemaes, C., Van Loey, A., and Hendrickx, M. (1998). Kinetics for isobaric–isothermal degradation of L-ascorbic acid. *J. Agric. Food Chem.* 46, 2001–2006. doi: 10.1021/jf9708251
- Van Wuytswinkel, O., Vansuyt, G., Grignon, N., Fourcroy, P., and Briat, J.-F. (1999). Iron homeostasis alteration in transgenic tobacco overexpressing ferritin. *Plant J.* 17, 93–97. doi: 10.1046/j.1365-313X.1999.00349.x
- Varotto, C., Maiwald, D., Pesaresi, P., Jahns, P., Salamini, F., and Leister, D. (2002). The metal ion transporter IRT1 is necessary for iron homeostasis and efficient photosynthesis in *Arabidopsis thaliana*. *Plant J.* 31, 589–599. doi: 10.1046/j.1365-313X.2002.01381.x
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Vasconcelos, M., Datta, K., Oliva, N., Khalekuzzaman, M., Torrizo, L., Krishnan, S., et al. (2003). Enhanced iron and zinc accumulation in transgenic rice with the ferritin gene. *Plant Sci.* 164, 371–378. doi: 10.1016/S0168-9452(02)00421-1
- Vasconcelos, M., Eckert, H., Arahana, V., Graef, G., Grusak, M., and Clemente, T. (2006). Molecular and phenotypic characterization of transgenic soybean expressing the *Arabidopsis* ferric chelate reductase gene, FRO2. *Planta* 224, 1116–1128. doi: 10.1007/s00425-006-0293-1
- Vert, G. A., Briat, J.-F., and Curie, C. (2003). Dual regulation of the *Arabidopsis* high-affinity root iron uptake system by local and long-distance signals. *Plant Physiol.* 132, 796–804. doi: 10.1104/pp.102.016089
- Vert, G., Barberon, M., Zelazny, E., Séguéla, M., Briat, J.-F., and Curie, C. (2009). *Arabidopsis* IRT2 cooperates with the high-affinity iron uptake system to

- maintain iron homeostasis in root epidermal cells. *Planta* 229, 1171–1179. doi: 10.1007/s00425-009-0904-8
- Vert, G., Grotz, N., Dédaldéchamp, F., Gaymard, F., Guerinot, M. L., Briat, J.-F., et al. (2002). IRT1, and Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* 14, 1223–1233. doi: 10.1105/tpc.001388
- von Wirén, N., Khodr, H., and Hider, R. C. (2000). Hydroxylated phytosiderophore species possess an enhanced chelate stability and affinity for iron(III). *Plant Physiol.* 124, 1149–1158. doi: 10.1104/pp.124.3.1149
- von Wirén, N., Klair, S., Bansal, S., Briat, J.-F., Khodr, H., Shioiri, T., et al. (1999). Nicotianamine chelates both Fe^{III} and Fe^{II}. Implications for metal transport in plants. *Plant Physiol.* 119, 1107–1114. doi: 10.1104/pp.119.3.1107
- Vucenik, I., and Shamsuddin, A. M. (2003). Cancer inhibition by inositol hexaphosphate (IP6) and inositol: from laboratory to clinic. *J. Nutr.* 133, 3778S–3784S.
- Walker, E. L., and Connolly, E. L. (2008). Time to pump iron: iron-deficiency-signaling mechanisms of higher plants. *Curr. Opin. Plant Biol.* 11, 530–535. doi: 10.1016/j.pbi.2008.06.013
- Walters, G. O., Jacobs, A., Worwood, M., Trevett, D., and Thomson, W. (1975). Iron absorption in normal subjects and patients with idiopathic haemochromatosis: relationship with serum ferritin concentration. *Gut* 16, 188–192. doi: 10.1136/gut.16.3.188
- Wang, M., Gruissem, W., and Bhullar, N. (2013). Nicotianamine synthase overexpression positively modulates iron homeostasis-related genes in high iron rice. *Front. Plant Sci.* 4:156. doi: 10.3389/fpls.2013.00156
- Waters, B. M., and Troupe, G. C. (2012). Natural variation in iron use efficiency and mineral remobilization in cucumber (*Cucumis sativus*). *Plant Soil* 352, 185–197. doi: 10.1007/s11104-011-0988-3
- Waters, B. M., Blevins, D. G., and Eide, D. J. (2002). Characterization of FRO1, a pea ferric-chelate reductase involved in root iron acquisition. *Plant Physiol.* 129, 85–94. doi: 10.1104/pp.010829
- Wells, H. F. (2016). *Vegetables and Pulses Yearbook Data*. USDA, USA.
- Wheby, M. S., Jones, L. G., and Crosby, W. H. (1964). Studies on iron absorption. Intestinal regulatory mechanisms. *J. Clin. Invest.* 43:1433. doi: 10.1172/JCI105019
- WHO (2001). *Iron Deficiency Anaemia: Assessment, Prevention and Control: A Guide for Programme Managers*. Geneva: World Health Organization.
- WHO (2008). *Worldwide Prevalence of Anaemia 1993-2005*. World Health Organisation.
- WHO (2016). *Biofortification of Staple Crops*.
- WHO and FAO (2006). *Guidelines on Food Fortification with Micronutrients*.
- Wirth, J., Poletti, S., Aeschlimann, B., Yakandawala, N., Drosse, B., Osorio, S., et al. (2009). Rice endosperm iron biofortification by targeted and synergistic action of nicotianamine synthase and ferritin. *Plant Biotechnol. J.* 7, 631–644. doi: 10.1111/j.1467-7652.2009.00430.x
- Wood, J. A., and Grusak, M. A. (2007). “Nutritional value of chickpea,” in *Chickpea Breeding and Management*, eds S. S. Yadav, R. J. Redden, W. Chen, and B. Sharma (Trowbridge: CAB International), 101–142. doi: 10.1079/9781845932138.005
- Wood, J. A., Knights, E. J., and Choct, M. (2011). Morphology of chickpea seeds (*Cicer arietinum* L.): comparison of desi and kabuli types. *Int. J. Plant Sci.* 172, 632–643. doi: 10.1086/659456
- Xu, Y., An, D., Liu, D., Zhang, A., Xu, H., and Li, B. (2012). Molecular mapping of QTLs for grain zinc, iron and protein concentration of wheat across two environments. *Field Crops Res.* 138, 57–62. doi: 10.1016/j.fcr.2012.09.017
- Yadav, S. S., Longnecker, N., Dusunceli, F., Bejiga, G., Yadav, M., Rizvi, A. H., et al. (2007). “Uses, consumption and utilisation,” in *Chickpea Breeding and Management*, eds S. Yadav, R. J. Redden, W. Chen, and B. Sharma (Trowbridge: CAB International), 72–100.
- Zhang, Y., Shi, R., Rezaul, K. M., Zhang, F., and Zou, C. (2010). Iron and zinc concentrations in grain and flour of winter wheat as affected by foliar application. *J. Agric. Food Chem.* 58, 12268–12274. doi: 10.1021/jf103039k
- Zheng, L., Cheng, Z., Ai, C., Jiang, X., Bei, X., Zheng, Y., et al. (2010). Nicotianamine, a novel enhancer of rice iron bioavailability to humans. *PLoS ONE* 5:e10190. doi: 10.1371/journal.pone.0010190

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Tan, Das Bhowmik, Hoang, Karbaschi, Johnson, Williams and Mundree. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Development of Molecular Markers for Iron Metabolism Related Genes in Lentil and Their Expression Analysis under Excess Iron Stress

Debjyoti Sen Gupta^{1,2*}, Kevin McPhee^{1*} and Shiv Kumar³

¹ Department of Plant Sciences, North Dakota State University, Fargo, ND, USA, ² Division of Crop Improvement, ICAR-Indian Institute of Pulses Research, Kanpur, India, ³ BIGM Program, International Center for Agricultural Research in the Dry Areas (ICARDA), Rabat-Institutes, Rabat, Morocco

OPEN ACCESS

Edited by:

Diego Rubiales,
Instituto de Agricultura Sostenible
(CSIC), Spain

Reviewed by:

Marcelino Perez De La Vega,
Universidad de León, Spain
Carla S. Santos,
Universidade Católica Portuguesa,
Portugal

*Correspondence:

Kevin McPhee
kevin.mcphee@ndsu.edu
Debjyoti Sen Gupta
debgpb@gmail.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 15 November 2016

Accepted: 30 March 2017

Published: 13 April 2017

Citation:

Sen Gupta D, McPhee K and
Kumar S (2017) Development
of Molecular Markers for Iron
Metabolism Related Genes in Lentil
and Their Expression Analysis under
Excess Iron Stress.
Front. Plant Sci. 8:579.
doi: 10.3389/fpls.2017.00579

Multiple genes and transcription factors are involved in the uptake and translocation of iron in plants from soil. The sequence information about iron uptake and translocation related genes is largely unknown in lentil (*Lens culinaris* Medik.). This study was designed to develop iron metabolism related molecular markers for *Ferritin-1*, *BHLH-1* (Basic helix loop helix), or FER-like transcription factor protein and *IRT-1* (Iron related transporter) genes using genome synteny with barrel medic (*Medicago truncatula*). The second objective of this study was to analyze differential gene expression under excess iron over time (2 h, 8 h, 24 h). Specific molecular markers were developed for iron metabolism related genes (*Ferritin-1*, *BHLH-1*, *IRT-1*) and validated in lentil. Gene specific markers for *Ferritin-1* and *IRT-1* were used for quantitative PCR (qPCR) studies based on their amplification efficiency. Significant differential expression of *Ferritin-1* and *IRT-1* was observed under excess iron conditions through qPCR based gene expression analysis. Regulation of iron uptake and translocation in lentil needs further characterization. Greater emphasis should be given to development of conditions simulating field conditions under external iron supply and considering adult plant physiology.

Keywords: lentil, gene expression, iron metabolism, qPCR expression analysis, molecular marker, ferritin

INTRODUCTION

Iron (Fe) uptake in plants is a complex physiological process governed by homeostatic mechanisms in the plant. Homeostatic mechanisms involve absorption, translocation and redistribution of Fe within the plant system at a particular concentration (10^{-9} – 10^{-4} mol/l) (Römheld and Schaaf, 2004). Lower iron concentration leads to Fe-deficiency symptoms including chlorosis and necrosis in leaves and ultimately loss in biomass as well as grain yield. Higher concentrations of Fe within the plant system results in generation of free radical species which damage various cellular components by interacting with protein, lipid, carbohydrates and even with DNA. According to Welch and Graham (2004), there are four different barriers controlling homeostatic mechanisms of mineral uptake in plants; (A) the root-soil interphase known as the rhizosphere, (B) root-cell plasma membrane, (C) translocation to edible plant organs (grains/tubers), and (D) bioavailability of minerals.

Ferritin is an iron-carrying protein in plants and has a multimeric (24-mer) cage-like structure that carries up to 4500 atoms of Fe within its core (Crichton et al., 1978; Wade et al., 1993). The

ferritin protein is highly conserved within the animal and plant kingdom (Ragland et al., 1990). Ferritin meets the metabolic need for iron when required by the metabolome as well as prevents any kind of oxidative stress (Harrison et al., 1998). Plant ferritin subunit sequences share between 39 and 49% similarity with mammalian ferritin sequences (Briat et al., 2009). This similarity increases when comparisons are made within the plant kingdom or among close plant families. Iron homeostasis is important due to the minute balance that exists between iron deficiency and toxicity and that affects plant physiology. Impaired plant physiology ultimately affects crop yield. Ferritin regulates iron homeostasis to prevent interaction of iron with other cellular components which may result in generation of free radicals during oxidative stress. In plants, ferritin consists of a single kind of subunit and ferritin bound Fe is highly bioavailable (Kalganekar and Lonnerdal, 2008).

Lentil (*Lens culinaris* Medik.) is a eudicot plant and uses strategy I where ferric iron is reduced at the rhizosphere and absorbed as ferrous iron by the root. Monocot plants use a different strategy to uptake iron from the soil (strategy II). The uptake of iron is mediated through phytosiderophores and the ferric iron enters the plant system through root in case of monocot plants (strategy II). In *Arabidopsis thaliana*, reduction of ferric Fe is accomplished by Fe reductase FRO2 (ferric reductase oxidase-2; Robinson et al., 1999). This was the first report of cloning and gene function elucidation of any major iron metabolism related gene in plants. Uptake of ferrous Fe into the root is carried out by the metal transporter IRT1 (iron-regulated transporter; Eide et al., 1996; Vert et al., 2002). The basic helix-loop-helix (BHLH) transcription factor family in plants is a ubiquitous regulator and is highly conserved, regulating different types of genes during transcription (Heim et al., 2003). The BHLH transcription factor or FIT (FER-like Fe deficiency-induced transcription factor) is reported to be responsible for high-level expression of FRO2 and IRT1 (Colangelo and Gueriot, 2004; Jakoby et al., 2004; Yuan et al., 2005). It is pertinent to mention that iron uptake, translocation and storage is a complex pathway and multiple genes or gene families are involved. However, from crop breeding point of view breeders always need a high-throughput and less time consuming techniques to identify few potential genotypes in a large set of germplasms. These three genes were targeted in lentil with a long-term objective in mind to develop an assay to find lentil genotypes which better perform under excess iron supply. The results of this experiment would give an initial thrust to such objectivities in lentil where limited amount of sequence information is available till date.

Development of gene specific markers and their utilization in understanding metabolic pathways are important genomic goals to achieve in any crop species for their effective utilization in genetic studies or molecular breeding applications *per se*. Availability of specific DNA markers for iron metabolism related genes in lentil are not available. The objectives of the study were to, (1) develop gene (*Ferritin-1*, *BHLH-1*, and *IRT-1*) specific molecular markers in lentil and (2) analyze their gene expression under excess iron over time.

MATERIALS AND METHODS

Plant Materials and Treatments

'CDC Redberry' (Vandenberg et al., 2006) seedlings were raised in the laboratory and fresh tissue was collected for DNA and RNA extraction. Seeds were germinated on wet filter paper in an incubator maintained at 25°C. Seedlings were transferred to 50 mL tubes containing distilled water for hydroponic growth with 16:8 h light: dark cycle and at 25°C for eight days after germination. After complete development of the first true leaf (of growth), treatments were applied 18–21 days after germination and included: (1) iron deficient condition (control with distilled water), (2) excess iron condition (addition of 500 µM of Fe-EDTA, 150 mM of sodium citrate, and 75 µM FeSO₄) (Lobréaux et al., 1995). Treatments were applied for 24 h and samples were collected 2, 8, and 24 h after treatment. Three biological replications were included for each treatment.

Development of Markers

Full length coding sequences (CDS) for three ferritin genes (*ferritin-1*, *ferritin-2*, *ferritin-3*) for *Medicago truncatula* were acquired from the NCBI (National Center for Biological Information) nucleotide database on 15 April 2015. The complete coding sequence of *Ferritin-2* mRNA (NCBI reference sequence: XM_003616637.1) of *M. truncatula* was downloaded in FASTA format and used to perform a nucleotide BLAST search against CDC Redberry 454 contig sequences in the Knowpulse database¹. The contig sequence with the highest bit score and lowest e-value and, therefore, having the highest similarity with the query sequence (*M. truncatula Ferritin-2*) was identified. The contig sequence was downloaded from the Knowpulse database and Primer-BLAST² was used to design primer pairs using default parameters (Table 1). One primer pair (FerrClo5) used for the development of qPCR compatible primers for the *Ferritin* gene in lentil. In addition, one primer pair specific to a lentil BHLH (Basic Helix Loop Helix) transcription factor or FER-like transcription factor gene sequence (Sen Gupta et al., 2016) was synthesized. Primers were also designed for the iron-related transporter gene based on the *IRT1* mRNA coding sequence (CDS) (LegumeIP database reference no. IMGAM[Medtr8g105030.1] of *M. truncatula* for the amplification of lentil *IRT-1* in the qPCR experiment. The amplicon of the ferritin gene as well as the BHLH transcription factor gene were beyond the range of optimum product size (>250 bp) for qPCR experiments and thus were gel purified using a gel purification kit (IBI, MIDSCI, St. Louis, MO, USA) (Vogelstein and Gillespie, 1979) following manufacturer's instructions and sequenced using the Sanger sequencing method (Etonbiosciences Inc., San Diego, CA, USA). The gene sequences were aligned with the respective *M. truncatula* mRNA sequences (*Ferritin-2* and *BHLH* transcription factor gene, respectively) and primers were designed for qPCR experiments based on the putative exonic sequences, their sequence identity, gap, and the desired product size using Primer3 software². Based on these sequences one primer pair for *Ferritin-1* and another primer pair

¹<http://knowpulse.usask.ca/portal/blast/nucleotide/nucleotide>

²<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>

TABLE 1 | Nucleotide BLAST results of *Medicago truncatula ferritin-2* gene sequence (NCBI reference no. XM_003616637.1) with CDC Redberry 454 contig sequences in Knowpulse database showing bit score, percent identity, and e-value (<http://knowpulse.usask.ca>).

Hit*	Bit score	Identity%	E-value
LcRBContig00605	700	91	0.00e+0
LcRBContig02360	530	90	1.53e-103
LcRBContig20139	142	93	4.44e-5
LcRBContig24460	167	94	1.39e-40
LcRBContig24460	167	94	1.39e-40
LcRBContig13391	167	94	1.39e-40
LcRBContig07868	167	94	1.39e-40
LcRBContig07177	167	94	1.39e-40
LcRBContig01318	167	94	1.39e-40
LcRBContig24151	111	91	7.13e-24

*First 10 relevant hits are shown here.

for the *BHLH-1* transcription factor were designed for qPCR. Primers for *IRT-1* were directly used in qPCR and were within the qPCR compatible product size range (<100 bp amplicon size).

Isolation of RNA and Synthesis of Complementary DNA

Total RNA was extracted from 100 mg of fresh leaves of individual treatments using the QIAGEN® RNeasy Mini Kit (QIAGEN, Valencia, CA, USA) according to manufacturer instructions. The quality of the RNA extracts were determined by the spectrophotometer Nano-Drop (ND-1000) (NanoDrop Technologies, Wilmington, DE, USA). To check the integrity of the RNA, the samples were stained, separated and visualized by electrophoresis in a 1% agarose gel. Details about the quality of the RNA samples can be found in Supplementary Table S1. The first strand of cDNA was synthesized from 1 µg of total RNA in a 20 µL reaction using SuperScript III First Strand Synthesis Supermix RT-PCR Kit (Invitrogen, USA). The cDNAs were diluted to 2 ng µL⁻¹.

Quantitative PCR

Three primer pairs were used for gene expression analysis, *Ferritin1* (developed using PCR based cloning and sequencing), *BHLH1* (developed using PCR based cloning and sequencing) and *IRT1* (primer designed based on *M. truncatula IRT1* gene sequence). Expression levels of mRNA were evaluated in a SYBR Green dye using an Applied Biosystems 7500 Fast Real-Time PCR System (Applied Biosystems, USA). PCR amplifications were carried out in triplicate in 20 µL reactions containing Maxima SYBR Green mixer (Fermentas, USA), 250 nM of each primer and 4 ng of cDNA. On each plate, the reference genes (*GADPH* and *Actin*) and negative controls were included. Amplification conditions were 50°C for 2 min, 95°C for 10 min, 40 cycles at 95°C for 15 s, 60°C for 1 min. The calibration curves for each primer pair were plotted using five serial dilutions of the cDNA in water. To verify the specificity of amplification a dissociation curve analysis step was added to the qPCR amplification protocol. Amplification efficiency, slope and R² value were determined for

each primer pair. Amplification efficiencies were calculated as $E = (10^{-1/\text{slope}} - 1) \times 100$.

Statistical Analysis of Gene Expression Analysis

Cycle threshold (C_T) values were determined using SDS software (Applied Biosystems, USA). Gene expression data were analyzed using the C_T values and amplification efficiency values using method 2^{-ΔΔCT} (Livak and Schmittgen, 2001). Geometric means of reference genes were used to normalize the C_T values of the individual samples. The program REST 2009—Relative Expression Software Tool (Pfaffl, 2001) was used to determine if the differences between the treatments were statistically significant ($P < 0.05$).

RESULTS

Development of Markers for *Ferritin-1*, *BHLH-1*, and *IRT-1* Genes

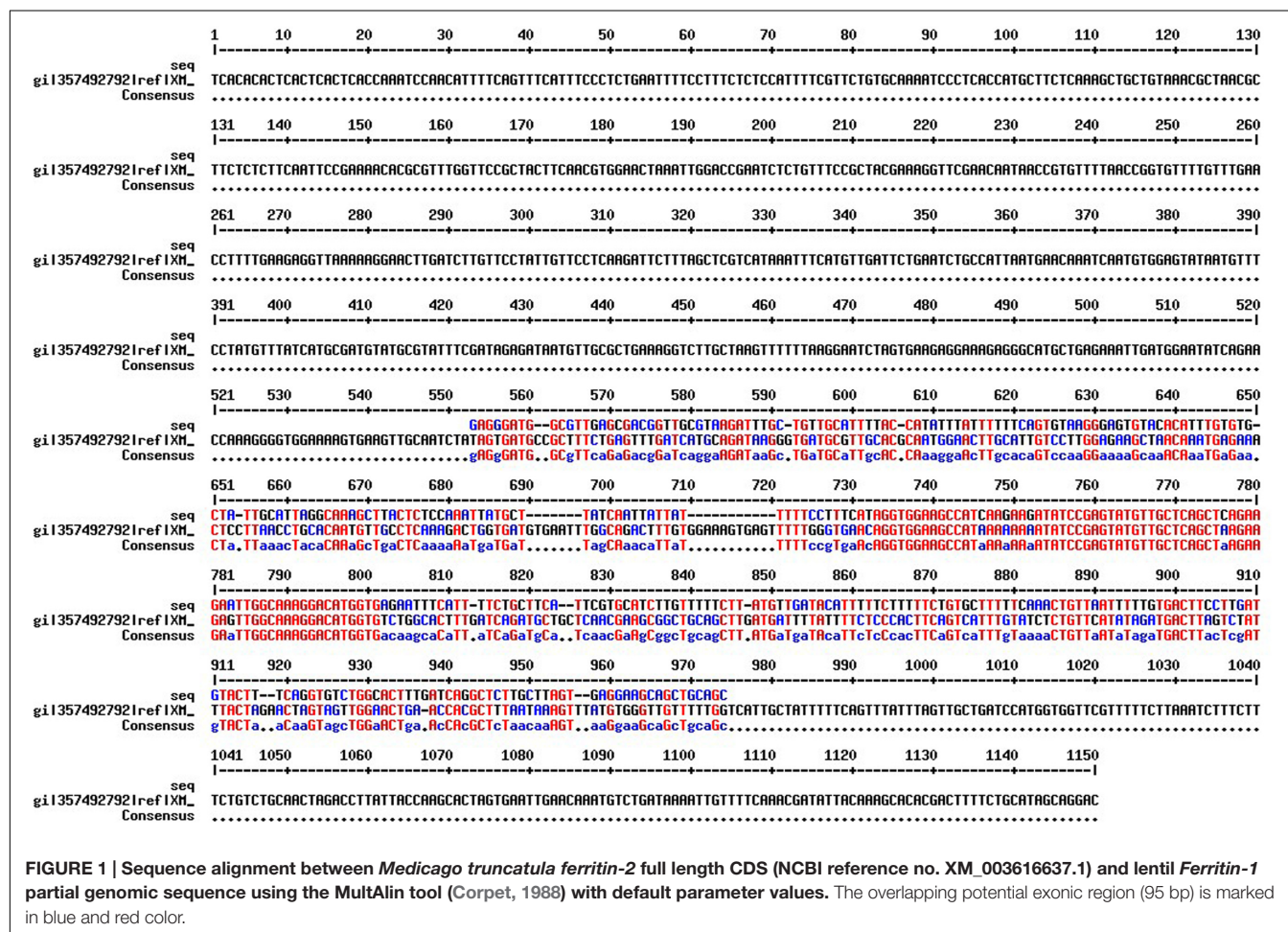
After performing BLASTn analysis using *ferritin-2* mRNA sequence of *Medicago truncatula* in the KnowPulse database (University of Saskatchewan, Canada) one contig sequence was identified, LcRBContig00605, based on BIT score (700), sequence identity (91%), and e-value (0) (Table 1). BLASTn results using other plant species resulted in identification of this contig sequence (LcRBContig00605) (data not shown). Optimum PCR conditions for the designed primers (FerrClo5) in an ABI 7500 thermocycler were established: 94°C for 5 minutes, followed by 30 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min followed by a final elongation step of 72°C for 5 min. The amplified DNA fragment was gel purified and sequenced using Sanger's method to obtain a 390 bp sequence. Alignment of the partial genomic DNA sequence with the *M. truncatula ferritin-2* mRNA sequence (NCBI reference sequence: XM_003616637.1) showed a 92 bp sequence overlap with no gap (Figure 1). This potential exonic sequence was used to design primers (*Ferritin-1*) using Primer-BLAST (Table 2).

Primer pairs developed in a previous study (Sen Gupta et al., 2016) were used to amplify the *BHLH-1* gene in CDC Redberry genomic DNA. Optimum PCR conditions for *BHLH-1* primer pairs in an ABI 7500 thermocycler were established: 94°C for 5 min, followed by 30 cycles of 94°C for 1 min, 60°C for 1 min, 72°C for 1 min followed by a final elongation step of 72°C for 5 min. The amplified fragment sequenced by Sanger's sequencing method and a 490 bp sequence was obtained. This sequence was aligned with *M. truncatula BHLH* mRNA sequence (NCBI reference number XM_003606283.1) and based on the alignment (Figure 2) a 75 bp sequence with no gap (potential exonic sequence) was used to design qPCR compatible primers for *BHLH-1* in lentil using Primer-BLAST.

Using a *M. truncatula* iron regulated transporter gene mRNA sequence [LegumeIP database reference no. IMGAMedtr8g105030.1] primer pairs (*IRT1*) were designed for the qPCR study.

TABLE 2 | Sequence and T_m (melting temperature) for primers designed based on the CDC Redberry contig (LcRBContig00605) for the *Ferritin-1* gene in lentil.

Primer name	Forward sequence (5'–3')	Reverse sequence (5'–3')	T_m (°C)
FerrClo1	TGCTGATAAGGGTGATGCGCT	GGCTTCCACCTGTTCAACCA	64
FerrClo2	AACCTGCACAGTGTTCGCTC	AGTGCCAGACACCATGTCCT	62
FerrClo3	GTTGCGCTGAAAGGCTTGTCT	GCCAAGTGCACATCACCAGT	62
FerrClo4	ACGTTGCGCTGAAAGGCTTGT	TGCCAAGTGCACATCACCAG	62
FerrClo5	CTGGTGATGTGCACTTGGCA	GCTGCAGCTGCTTCCTCACT	62



Dissociation curve analysis of the three pairs of primers (*Ferritin-1*, *BHLH-1*, *IRT-1*) showed specific amplification (Figure 3). Amplification efficiency of the designed primer pairs and reference genes (*GADPH*, *Actin*) were found to be >90% with the exception of *BHLH-1* primer pairs (Table 3). Slope values ranged from -0.02 to -3.55 and R^2 values ranged between 0.0034 and 0.9972.

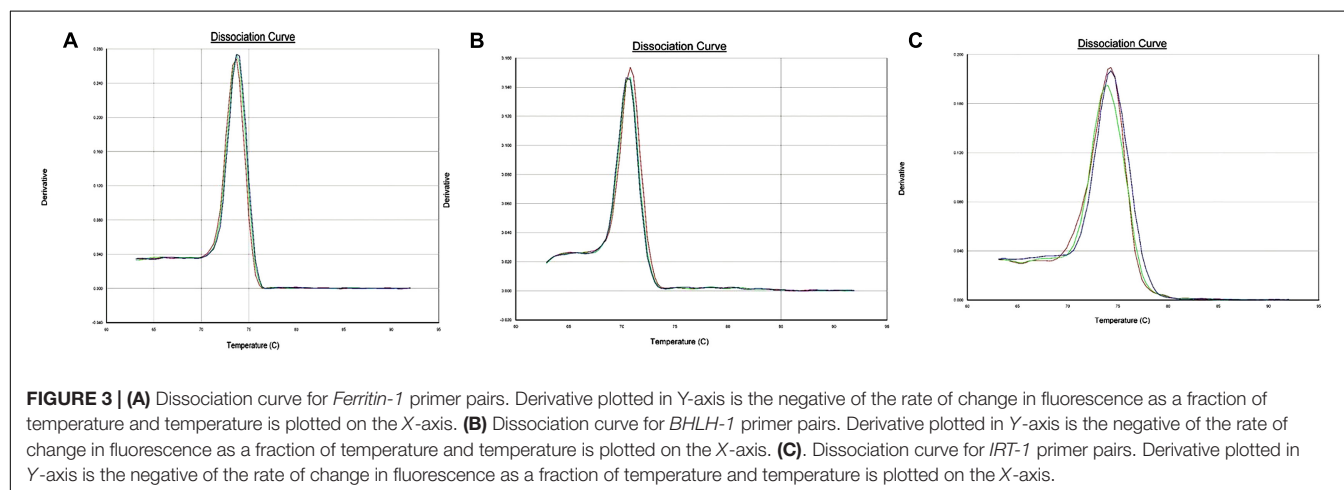
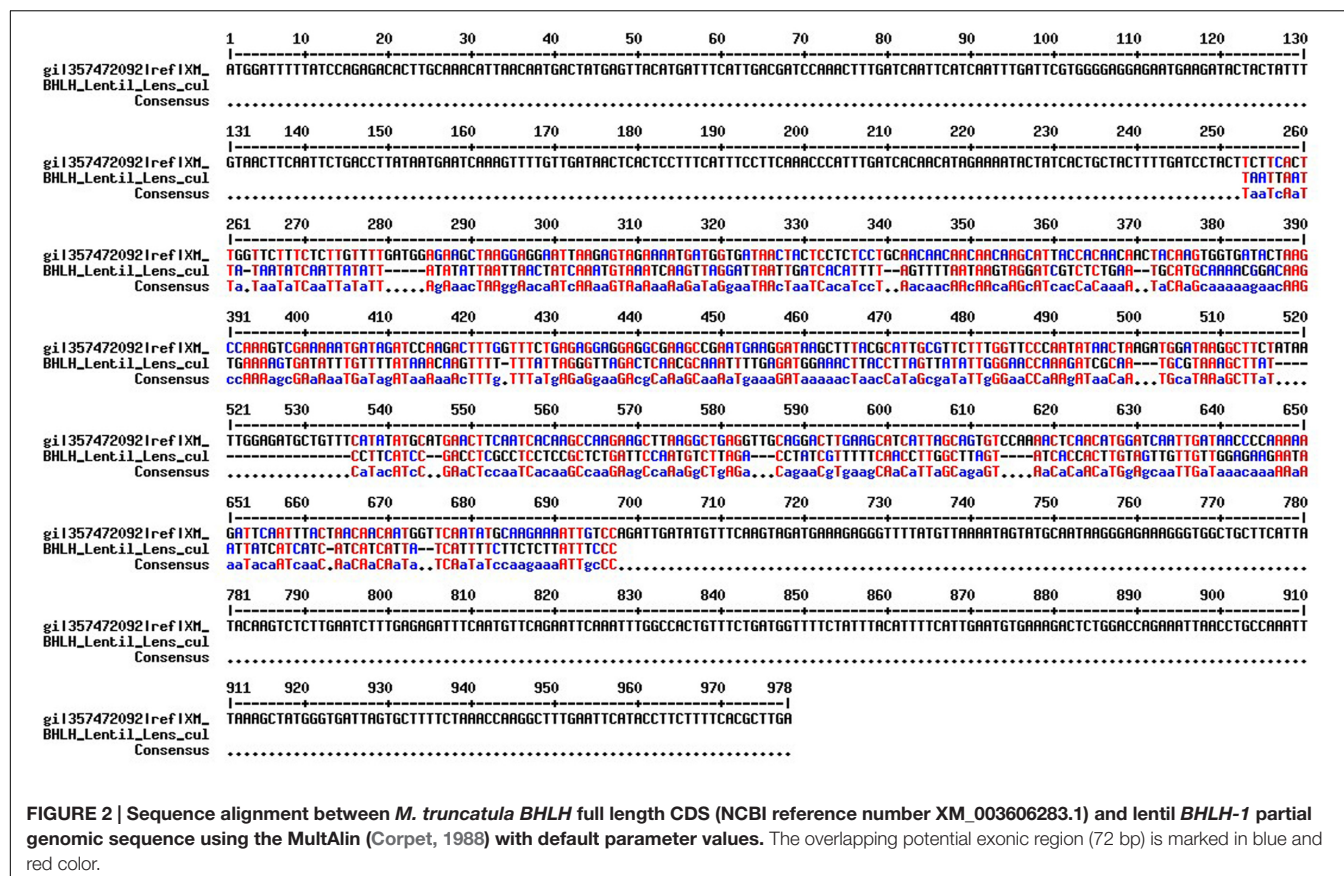
Expression Analysis of *Ferritin-1* and *IRT-1* Genes

Using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001), changes in gene transcripts were calculated for the treated samples (under excess iron condition) compared to the control

treatments (iron-deficient condition) (Table 4). The changes in gene transcript levels for *Ferritin-1* and *IRT-1* genes were not significantly different for the shoot tissue (Table 5). A 2.72-fold increase in *Ferritin-1* gene transcripts was observed in root tissue after 2 h of iron treatment ($P < 0.05$) (Table 5). Similarly, a 3.6-fold increase in *IRT-1* gene transcripts was observed ($P < 0.05$) (Table 5).

DISCUSSION

Iron uptake from the soil and translocation within the plant is a complex physiological process. It involves multiple genes and transcription factors. The magnitude of mRNA transcript



synthesis under excess iron conditions for iron metabolism related genes (*Ferritin-1*, *BHLH-1*, *IRT-1*) in lentil was evaluated in this study. Two genes, *Ferritin-1* and *IRT-1*, were quantitatively assayed for differential gene expression as they exhibited amplification efficiency of > 90 percent (Udvardi et al., 2008).

Dissociation curve analysis (Figure 3) which is the dsDNA melting curve analysis (Udvardi et al., 2008) added at the end of PCR run showed the specificity for single amplicon amplification and expected melting temperature for the individual primer

pairs. All of the three primer pairs exhibited a typical single peak with expected melting temperatures (Figure 3). Gene expression quantification values (C_T values) were normalized using geometric means of C_T values of the two reference genes (*GADPH*, *Actin*) (Vandesompele et al., 2002). *Actin* and *GADPH* were used in studies in lentil, pea and common bean exhibiting stability of expression across tissues and plant parts (Saha and Vandemark, 2012, 2013, DeLaat et al., 2014). The objective behind the normalization of qPCR data was to remove the

TABLE 3 | Amplification statistics for one *Ferritin-1*, one *BHLH-1*, one *IRT-1* gene specific primer pairs, and one primer pair for each reference gene (*GADPH*, *Actin*).

Gene	Forward sequence	Reverse sequence	T_m (°C)	Size (bp)	Slope	R^2	E	Reference
<i>Ferritin-1</i>	AGATATCCGAGTATGTTGCTCAG	AAGATGCACGAATGAAGCAGAAA	61	84	-3.32	0.9968	100.07	Current work
<i>IRT-1</i>	GTCGCTGTTTTGCTAGGTGC	GTGAGCTTCTCCTCTTCCCT	61	159	-3.12	0.9954	109.18	Current work
<i>BHLH-1</i>	TTATTAGGGTTAGACTCAACGCA	TTGCGATCTTTGGTTCCCA	59	74	-0.02	0.0034	6.55e+42	Sen Gupta et al., 2016
<i>GADPH</i>	TGGGCGAAAACCTCCACTTTG	GAATTGCTGCAGCCTTGTGA	60	57	-3.15	0.9954	107.71	Saha and Vandemark, 2012
<i>Actin</i>	CCAAATCATGTTTGAGGCTTTTAA	GTGAAAGAACGGCCTGAATAGC	60	64	-3.55	0.9972	91.25	Saha and Vandemark, 2012

Here, T_m = melting temperature, Size = amplicon length, Slope = slope of the trend line in amplification efficiency graph, R^2 = regression coefficient, E = amplification efficiency.

TABLE 4 | Differentially expressed *Ferritin-1* and *IRT-1* genes in CDC Redberry shoot and root tissues over time (2, 8 and 24 h) in three replicates under excess iron.

Gene	Plant tissue	Time course								
		2 h	2 h	2 h	8 h	8 h	8 h	24 h	24 h	24 h
<i>Ferritin-1</i>	Shoot tissue	0.29	1.0	3.03	0.47	0.79	1.45	2.7	1.41	0.83
		0.37	1.47	0.38	0.15	0.44	1.79	1.38	1.0	0.20
<i>IRT-1</i>	Root tissue	1.81	3.29	3.05	0.22	0.52	1.45	0.64	0.32	0.82
		1.70	5.03	4.59	0.30	0.55	1.73	0.73	0.44	1.14

TABLE 5 | Significance of differential expression of samples over time (TC) in excess iron in relation to control samples in shoot and root tissue of CDC Redberry genotype.

TC (h)	Gene	Tissue	N	E	SE	95% CI	$P(H1)$	Remark
2 h	<i>Ferritin-1</i>	Shoot	3	0.474	0.241–0.938	0.159–1.264	0.199	NS
8 h	<i>Ferritin-1</i>	Shoot	3	1.056	0.763–1.411	0.711–1.843	0.724	NS
24 h	<i>Ferritin-1</i>	Shoot	3	1.049	0.589–2.150	0.377–2.792	0.832	NS
2 h	<i>Ferritin-1</i>	Root	3	2.724	1.866–4.644	1.342–5.267	0	Up regulated, significant
8 h	<i>Ferritin-1</i>	Root	3	0.554	0.310–1.018	0.228–1.395	0.28	NS
24 h	<i>Ferritin-1</i>	Root	3	0.558	0.383–0.796	0.096		NS
2 h	<i>IRT-1</i>	Shoot	3	0.591	0.223–1.443	0.116–2.883	0.539	NS
8 h	<i>IRT-1</i>	Shoot	3	0.487	0.218–1.634	0.162–1.835	0.298	NS
24 h	<i>IRT-1</i>	Shoot	3	0.65	0.283–1.280	0.211–2.734	0.517	NS
2 h	<i>IRT-1</i>	Root	3	3.563	2.186–5.066	1.874–5.405	0	Up regulated, significant
8 h	<i>IRT-1</i>	Root	3	0.672	0.386–1.101	0.313–1.640	0.245	NS

Here, N = number of biological replications, E = Differential expression, SE = standard error, $P(H1)$ = Probability of alternative hypothesis.

sampling error, which may arise due to RNA quantity and quality differences across samples.

In this study, we developed gene-specific molecular markers for three genes (*Ferritin-1*, *BHLH-1*, *IRT-1*) in lentil. Primers for *Ferritin-1* and *IRT-1* were used in differential gene expression analysis. Partial genomic DNA sequences of *Ferritin-1* and *BHLH-1* were submitted to the NCBI database. These sequences are available to clone full length genomic sequences of each gene in lentil. The partial genomic DNA sequence *BHLH-1* gene can be further analyzed and used to develop qPCR compatible primers for this gene. It can be hypothesized from the comparative genomic synteny of lentil with *M. truncatula* (Phan et al., 2007)

that a ferritin gene family does exist in lentil and other ferritin genes in *M. truncatula* (*ferritin-1* and *ferritin-3*) could be used to develop molecular markers for the respective ferritin genes in lentil. In addition, once the lentil whole genome sequence is released cloning and characterization of ferritin and other iron metabolism related genes will be easier.

In gene expression analysis under excess iron it was observed that only samples with 2 h excess iron treatments exhibited significant differential gene expression (Table 5) for both genes (*Ferritin-1* and *IRT-1*) in root tissues. The absence of such kinetics in gene expression change for samples that were given 8 or 24 h excess iron treatments across the tissues was observed.

The possible reason could be the different iron homeostasis mechanisms in lentil compared to other plant species studied under similar conditions. Development of an assay to find out the reason behind such variation could first start with the standardization of external iron treatments in lentil. In common bean by applying identical excess iron concentration (Lobréaux et al., 1995) in leaf tissue similar kinetics of differential gene expression of ferritin genes (*PvFer1*, *PvFer2*, and *PvFer3*) were observed (DeLaat et al., 2014). Out of the three genotypes (IAC-Diplomata, Carioca, and BAT 477) used there had been significant genotypic differences of ferritin gene expression for two ferritin genes (*PvFer1*, *PvFer2*) (DeLaat et al., 2014). There were no significant differences among the treatments (control with distilled water, osmotic shock causing polyethylene glycol (PEG) treated, excess iron treated, PEG + excess iron treated) for any of the ferritin genes (DeLaat et al., 2014). The interaction between time and treatment was only significant for the *PvFer2* and interaction between time and cultivar was significant for the *PvFer3* ferritin gene (DeLaat et al., 2014). In most of the treatments ferritin genes were up regulated, however, there were treatments where *PvFer1* and *PvFer3* were down regulated (DeLaat et al., 2014) over time. The above-mentioned facts for common bean ferritin genes support the results we obtained in the case of *Ferritin-1* and *IRT-1* genes under identical conditions. Further, the gene expression levels for iron metabolism related genes were low in lentil as evident by the high C_T values. Number of biological replications may be increased to improve power of the test. The difference between seedling and adult plant physiology should be taken into consideration in future experiments. In summary, gene specific markers were developed for three iron metabolism related genes (*Ferritin-1*, *BHLH-1*, *IRT-1*) in lentil using PCR based cloning and

significant differential expression was observed for *Ferritin-1* and *IRT-1* genes at the transcriptional level.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: KM, DSG, SK. Performed the experiments: DSG. Analyzed the data: DSG. Contributed reagents/materials/analysis tools: KM. Wrote the paper: DSG, KM, SK. All authors have read and approved the manuscript.

FUNDING

This work was supported by the North Dakota State University and a doctoral fellowship was awarded to DSG from Indian Council of Agricultural Research.

ACKNOWLEDGMENTS

DSG thanks Indian Council of Agricultural Research, New Delhi for awarding Netaji Subhas ICAR International Fellowship for Doctoral study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00579/full#supplementary-material>

REFERENCES

- Briat, J. F., Céline, D., Karl, R., and Frédéric, G. (2009). Ferritins and iron storage in plants. *Biochim. Biophys. Acta*. 1800, 806–814. doi: 10.1016/j.bbagen.2009.12.003
- Colangelo, E. P., and Guerinot, M. L. (2004). Put the metal to the petal: metal uptake and transport throughout plants. *Curr. Opin. Plant Biol.* 9, 322–330. doi: 10.1016/j.pbi.2006.03.015
- Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* 16, 10881–10890. doi: 10.1093/nar/16.22.10881
- Crichton, R. R., Ponce-Ortiz, Y., Koch, M. H., Parfait, R., and Stuhmann, H. B. (1978). Isolation and characterization of phytoferritin from pea (*Pisum sativum*) and lentil (*Lens esculenta*). *Biochem. J.* 171, 349–356. doi: 10.1042/bj1710349
- DeLaat, D. M., Colombo, C. A., Chiorato, A. F., and Carbonell, S. A. M. (2014). Induction of ferritin synthesis by water deficit and iron excess in common bean (*Phaseolus vulgaris* L.). *Mol. Biol. Rep.* 41, 1427–1435. doi: 10.1007/s11033-013-2987-3
- Eide, D., Broderius, M., Fett, J., and Guerinot, M. L. (1996). A novel iron-regulated metal transporter from plants identified by functional expression in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5624–5628. doi: 10.1073/pnas.93.11.5624
- Harrison, P. M., Hempstead, P. D., Artymiuk, P. J., and Andrews, S. C. (1998). Structure–function relationships in the ferritins. *Met. Ions Biol. Syst.* 35, 435–477.
- Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. C. (2003). The basic helix–loop–helix transcription factor family in plants: a genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* 20, 735–747. doi: 10.1093/molbev/msg088
- Jakoby, M., Wang, H. Y., Reidt, W., Weisshaar, B., and Bauer, P. (2004). FRU (BHLH029) is required for induction of iron mobilization genes in *Arabidopsis thaliana*. *FEBS Lett.* 577, 528–534. doi: 10.1016/j.febslet.2004.10.062
- Kalgaonkar, S., and Lonnerdal, B. (2008). Effects of dietary factors on iron uptake from ferritin by Caco-2 cells. *J. Nutr. Biochem.* 19, 33–39. doi: 10.1016/j.jnutbio.2007.02.001
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) method. *Methods* 4, 402–408. doi: 10.1006/meth.2001.1262
- Lobréaux, S., Thoirion, S., and Briat, J. F. (1995). Induction of ferritin synthesis in maize leaves by an iron-mediated oxidative stress. *Plant J.* 8, 443–449. doi: 10.1046/j.1365-313X.1995.08030443.x
- Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.* 29:e45. doi: 10.1093/nar/29.9.e45
- Phan, H. T., Ellwood, S. R., Hane, J. K., Ford, R., Materne, M., and Oliver, R. P. (2007). Extensive macrosynteny between *Medicago truncatula* and ssp. *culinaris*. *Theor. Appl. Genet.* 114, 549–558. doi: 10.1007/s00122-006-0455-3
- Ragland, M., Briat, J. F., Gagnon, J., Laulhere, J. P., Massenet, O., and Theil, E. C. (1990). Evidence for conservation of ferritin sequences among plants and animals and for a transit peptide in soybean. *J. Biol. Chem.* 265, 18339–18344.
- Robinson, N. J., Procter, C. M., Connolly, E. L., and Guerinot, M. L. (1999). A ferric-chelate reductase for iron uptake from soils. *Nature* 397, 694–697. doi: 10.1038/17800
- Römhild, V., and Schaaf, G. (2004). Iron transport in plants: future research in view of a plant nutritionist and a molecular biologist. *Soil Sci. Plant Nutr.* 50, 1003–1012. doi: 10.1080/00380768.2004.10408567
- Saha, G. C., and Vandemark, G. J. (2012). Evaluation of expression stability of candidate reference genes among green and yellow pea cultivars (*Pisum*

- sativum* L.) subjected to abiotic and biotic stress. *Am. J. Plant Sci.* 3, 235–242. doi: 10.4236/ajps.2012.32028
- Saha, G. C., and Vandemark, G. J. (2013). Stability of expression of reference genes among different lentil (*Lens culinaris*) genotypes subjected to cold stress, white mold disease, and *Aphanomyces* root rot. *Plant Mol. Biol. Rep.* 31, 1109–1115. doi: 10.1007/s11105-013-0579-y
- Sen Gupta, D., Cheng, P., Sablok, G., Thavarajah, P., Coyne, C. J., Kumar, S., et al. (2016). Development of a panel of unigene-derived polymorphic EST-SSR markers in lentil using public database information. *Crop J.* 4, 425–433. doi: 10.1016/j.cj.2016.06.012
- Udvardi, M. K., Czechowski, T., and Scheible, W. R. (2008). Eleven golden rules of quantitative RT-PCR. *Plant Cell* 20, 1736–1737. doi: 10.1105/tpc.108.061143
- Vandenberg, A., Banniza, S., Warkentin, T. D., Ife, S., Barlow, B., McHale, S., et al. (2006). CDC Redberry lentil. *Can J. Plant Sci.* 86, 497–498. doi: 10.4141/P05-071
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van, R. N., De Paepe, A., et al. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3, RESEARCH0034. doi: 10.1186/gb-2002-3-7-research0034
- Vert, G., Grotz, N., Dédaldéchamp, F., Gaymard, F., Guerinot, M. L., Briat, J. F., et al. (2002). IRT1, an Arabidopsis transporter essential for iron uptake from the soil and for plant growth. *Plant Cell* 14, 1223–1233. doi: 10.1105/tpc.001388
- Vogelstein, B., and Gillespie, D. (1979). Preparative and analytical purification of DNA from agarose. *Proc. Natl. Acad. Sci. U.S.A.* 76, 615–619. doi: 10.1073/pnas.76.2.615
- Wade, V. J., Treffry, A., Laulhère, J. P., Bauminger, E. R., Cleton, M. I., Mann, S., et al. (1993). Structure and composition of ferritin cores from pea seed (*Pisum sativum*). *Biochim. Biophys. Acta* 1161, 91–96. doi: 10.1016/0167-4838(93)90201-2
- Welch, R. M., and Graham, R. D. (2004). Breeding for micronutrients in staple food crops from a human nutrition perspective. *J. Exp. Bot.* 55, 353–364. doi: 10.1093/jxb/erh064
- Yuan, Y. X., Zhang, J., Wang, D. W., and Ling, H. Q. (2005). AtbHLH29 of *Arabidopsis thaliana* is a functional ortholog of tomato FER involved in controlling iron acquisition in strategy I plants. *Cell Res* 15, 613–621. doi: 10.1038/sj.cr.7290331

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Sen Gupta, McPhee and Kumar. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Characterization of the Heme Pocket Structure and Ligand Binding Kinetics of Non-symbiotic Hemoglobins from the Model Legume *Lotus japonicus*

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica – Universidade Nova
de Lisboa, Portugal

Reviewed by:

José Alejandro Heredia-Guerrero,
Fondazione Istituto Italiano di
Tecnologia, Italy
Francisca Sevilla,
Centro de Edafología y Biología
Aplicada del Segura (CSIC), Spain

*Correspondence:

Cristiano Viappiani
cristiano.viappiani@fis.unipr.it
Manuel Becana
becana@eead.csic.es

† These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 30 November 2016

Accepted: 09 March 2017

Published: 04 April 2017

Citation:

Calvo-Begueria L, Cuypers B,
Van Doorslaer S, Abbruzzetti S,
Bruno S, Berghmans H, Dewilde S,
Ramos J, Viappiani C and Becana M
(2017) Characterization of the Heme
Pocket Structure and Ligand Binding
Kinetics of Non-symbiotic
Hemoglobins from the Model Legume
Lotus japonicus.
Front. Plant Sci. 8:407.
doi: 10.3389/fpls.2017.00407

Laura Calvo-Begueria^{1†}, Bert Cuypers^{2†}, Sabine Van Doorslaer², Stefania Abbruzzetti^{3,4}, Stefano Bruno⁵, Herald Berghmans⁶, Sylvia Dewilde⁶, Javier Ramos¹, Cristiano Viappiani^{4,7*} and Manuel Becana^{1*}

¹ Departamento de Nutrición Vegetal, Estación Experimental de Aula Dei, Consejo Superior de Investigaciones Científicas, Zaragoza, Spain, ² Department of Physics, University of Antwerp, Antwerp, Belgium, ³ Dipartimento di Bioscienze, Università degli Studi di Parma, Parma, Italy, ⁴ NEST, Istituto Nanoscienze, Consiglio Nazionale delle Ricerche, Pisa, Italy, ⁵ Dipartimento di Farmacia, Università degli Studi di Parma, Parma, Italy, ⁶ Department of Biomedical Sciences, University of Antwerp, Antwerp, Belgium, ⁷ Dipartimento di Fisica e Scienze della Terra, Università degli Studi di Parma, Parma, Italy

Plant hemoglobins (Hbs) are found in nodules of legumes and actinorhizal plants but also in non-symbiotic organs of monocots and dicots. Non-symbiotic Hbs (nsHbs) have been classified into two phylogenetic groups. Class 1 nsHbs show an extremely high O₂ affinity and are induced by hypoxia and nitric oxide (NO), whereas class 2 nsHbs have moderate O₂ affinity and are induced by cold and cytokinins. The functions of nsHbs are still unclear, but some of them rely on the capacity of hemes to bind diatomic ligands and catalyze the NO dioxygenase (NOD) reaction (oxyferrous Hb + NO → ferric Hb + nitrate). Moreover, NO may nitrosylate Cys residues of proteins. It is therefore important to determine the ligand binding properties of the hemes and the role of Cys residues. Here, we have addressed these issues with the two class 1 nsHbs (LjGlb1-1 and LjGlb1-2) and the single class 2 nsHb (LjGlb2) of *Lotus japonicus*, which is a model legume used to facilitate the transfer of genetic and biochemical information into crops. We have employed carbon monoxide (CO) as a model ligand and resonance Raman, laser flash photolysis, and stopped-flow spectroscopies to unveil major differences in the heme environments and ligand binding kinetics of the three proteins, which suggest non-redundant functions. In the deoxyferrous state, LjGlb1-1 is partially hexacoordinate, whereas LjGlb1-2 shows complete hexacoordination (behaving like class 2 nsHbs) and LjGlb2 is mostly pentacoordinate (unlike other class 2 nsHbs). LjGlb1-1 binds CO very strongly by stabilizing it through hydrogen bonding, but LjGlb1-2 and LjGlb2 show lower CO stabilization. The changes in CO stabilization would explain the different affinities of the three proteins for gaseous ligands. These affinities are determined by the dissociation rates and follow the order LjGlb1-1 > LjGlb1-2 > LjGlb2. Mutations LjGlb1-1 C78S and LjGlb1-2 C79S caused important alterations in protein dynamics and stability, indicating

a structural role of those Cys residues, whereas mutation LjGlb1-1 C8S had a smaller effect. The three proteins and their mutant derivatives exhibited similarly high rates of NO consumption, which were due to NOD activity of the hemes and not to nitrosylation of Cys residues.

Keywords: heme cavity, ligand binding, nitric oxide dioxygenase, non-symbiotic hemoglobins, *Lotus japonicus*

INTRODUCTION

The first plant hemoglobins (Hbs) were discovered in the root nodules of legumes and accordingly designated leghemoglobins (Appleby, 1984). The discovery of Hbs was subsequently extended not only to the nodules of *Parasponia* and actinorhizal plants (Tjepkema et al., 1986; Bogusz et al., 1988), but also to non-symbiotic tissues of monocots (Taylor et al., 1994; Arredondo-Peter et al., 1997), legumes (Andersson et al., 1996), and *Arabidopsis thaliana* (Trevaskis et al., 1997). Phylogenetic analyses showed that these non-symbiotic Hbs (nsHbs) belong to two distinct clades, termed class 1 and class 2 (Smagghe et al., 2009). Class 1 nsHbs have an extremely high O₂ affinity and are induced by hypoxia (Trevaskis et al., 1997; Smagghe et al., 2009) and by exposure to nitrate, nitrite, or nitric oxide (NO) (Sasakura et al., 2006). These proteins may play a role in plant survival by increasing the energy status of the cells under hypoxic conditions (Igamberdiev and Hill, 2004). The underlying molecular mechanism is thought to be the Hb/NO cycle, in which the NO dioxygenase (NOD) activity of Hb plays a critical role (Igamberdiev and Hill, 2004). In this reaction, the oxyferrous Hb dioxygenates NO to yield nitrate and ferric Hb. The NOD activities of a few class 1 nsHbs, including *A. thaliana* Hb1 (AtGlb1), have been measured *in vitro* (Perazzolli et al., 2004; Igamberdiev et al., 2006; Smagghe et al., 2008). However, AtGlb1 may be also nitrosylated by NO on Cys residues and this might affect its function (Perazzolli et al., 2004). Class 2 nsHbs have a moderate O₂ affinity and are induced by low temperature and cytokinins but not by hypoxia (Hunt et al., 2001).

At the molecular level, the control of hemeprotein function is tightly coupled to the structure and ligand-binding dynamics of the heme pocket. For plant Hbs, early studies on the heme properties have been focused on leghemoglobins (Appleby et al., 1976; Rousseau et al., 1983), although more recently some information has become available about the hemes of nsHbs (Das et al., 1999; Ioaniteanu et al., 2005; Abbruzzetti et al., 2007; Bruno et al., 2007a). Class 1 and class 2 Hbs and some animal globins, such as neuroglobin and cytoglobin, are hexacoordinate because, in the absence of exogenous ligands, they have the fifth (proximal) and sixth (distal) positions of the heme iron coordinated to His residues; in contrast, leghemoglobins and mammalian Hb and myoglobin (Mb) are pentacoordinate because the fifth position of the heme iron is coordinated to a His residue but the sixth position is open for ligand binding (Kakar et al., 2010). Interestingly, a system of hydrophobic cavities, capable of transiently stocking reactants and/or products, was proposed to be central to sustain the turnover of NOD activity in class 1 nsHbs (Spyrakakis et al., 2011; Abbruzzetti et al., 2013), as may occur for

Mb (Brunori, 2001) and neuroglobin (Uzan et al., 2004; Brunori et al., 2005).

Legumes are important in agriculture for two main reasons: they are a source of protein for animal and human nutrition, and they can establish nitrogen-fixing symbioses with soil rhizobia, allowing to minimize the supply of contaminant and costly nitrogen fertilizers. Besides leghemoglobins, whose expression is restricted to nodules, legumes contain nsHbs in leaves, roots, and nodules (Andersson et al., 1996; Bustos-Sanmamed et al., 2011). In fact, some of these nsHbs have important functions in the onset of symbiosis (Fukudome et al., 2016) and exhibit high expression in nodules compared with other plant organs (Bustos-Sanmamed et al., 2011). In this work, we have studied the nsHbs of *Lotus japonicus*, a model legume for classical and molecular genetics (Handberg and Stougaard, 1992). We have selected *L. japonicus* instead of *A. thaliana* as plant material because the information gained about the nsHbs of the former species will help define their role in symbiosis and will facilitate translational genomics to crop legumes. More specifically, the use of *L. japonicus* has allowed us to compare herein the biochemical properties of two class 1 nsHbs (LjGlb1-1 and LjGlb1-2) and a class 2 nsHb (LjGlb2) that greatly differ in their expression profiles in plant tissues (Bustos-Sanmamed et al., 2011) and in their O₂ affinities (Sainz et al., 2013). These previous results from one of our laboratories prompted us to investigate the heme environment properties of LjGlbs, as well as the contribution of their Cys residues to protein stability and ligand binding kinetics. To accomplish these objectives, we have performed a detailed spectroscopic study of the wild-type and the mutated proteins and have measured their NOD activities.

MATERIALS AND METHODS

Protein Purification and Identification of Disulfide Bond

The three nsHbs of *L. japonicus* were cloned into the Champion pET200/D-TOPO expression vector (Invitrogen) and expressed with an N-terminal poly-His tag in *Escherichia coli* BL21 Star (DE3) cells (Invitrogen) or C41(DE3) cells (Lucigen; Middleton, MI, USA) following conventional protocols (Sainz et al., 2013). The mutated versions of LjGlb1-1 C8S, LjGlb1-1 C78S, LjGlb1-2 C79S, and LjGlb2 C65S were obtained by PCR-based single-site substitutions using appropriate primers (Mutagenex; Somerset, NJ, USA). The DNA constructs were entirely sequenced and the amino acid substitutions (Supplementary Figure S1) were verified by matrix-assisted laser desorption and ionization time-of-flight (MALDI-TOF) mass spectrometry analysis of the trypsinized protein using a 4800 TOF/TOF instrument (AB

Sciex; Framingham, MA, USA). The proteins were purified using ammonium sulfate fractionation, metal-affinity chromatography, and anion-exchange chromatography, as reported earlier (Sainz et al., 2013). Purification of LjGlb1-1 was carried out in the presence of 200 mM NaCl to avoid precipitation of the dimeric form. The presence of LjGlb1-1 homodimer was demonstrated by fast protein liquid chromatography (FPLC) and mass spectrometry. For FPLC, the purified protein (10 mg) was loaded on a gel filtration column (Superdex 200 HR 10/30) coupled to an ÄKTA FPLC chromatography system (GE Healthcare Life Sciences), and was eluted with 50 mM potassium phosphate (pH 7.0) containing 150 mM NaCl at a flow rate of 0.5 ml min^{-1} . The void volume was calculated with dextran blue (0.1 mg ml^{-1}) and the column was calibrated with cytochrome *c* (12.4 kDa), Mb (17 kDa), ovalbumin (44.3 kDa), and bovine serum albumin (66 kDa). For molecular mass determinations, the purified protein was diluted 1:50 with 0.2% trifluoroacetic acid and analyzed by MALDI-TOF mass spectrometry. Calibration was performed with a mixture of albumin, trypsinogen, and protein A (mass range between 22,306 and 66,431 Da), and the accuracy was $\pm 50 \text{ Da}$ at 40 kDa.

Resonance Raman Spectroscopy

Resonance Raman (RR) spectra were acquired using a Dilor XY-800 spectrometer in low-dispersion mode using a liquid N_2 -cooled CCD detector. The excitation source was a Spectra Physics (Mountain View, CA, USA) BeamLok 2060 Kr^+ laser operating at 413.1 nm. The spectra were recorded at room temperature and the protein solutions were magnetically stirred at 500 rpm in order to avoid local heating and photochemical decomposition. The slit width used during the experiments was 200 μm . In general, 12–15 spectra were acquired with an integration time of 150–180 s each. Spikes due to cosmic rays were removed by omitting the highest and lowest data points for each frequency and by averaging the remaining values. Typical sample concentrations were in the order of 40–60 μM .

Ligand Binding Kinetics

Laser flash photolysis (LFP) experiments were performed at 20°C using a laser photolysis system (Edinburgh Instruments LP920, UK) equipped with a frequency-doubled, Q-switched Nd:YAG laser (Quanta-Ray, Spectra Physics). The carbon monoxide (CO)-ferrous Hb complexes were prepared in sealed $4 \times 10 \text{ mm}$ quartz cuvettes with 1 ml of 100 mM potassium phosphate buffer (pH 7.0) containing 1 mM EDTA. In the case of LjGlb1-1 this buffer was supplemented with 200 mM NaCl to improve protein stability. The buffer was equilibrated with mixtures of CO and N_2 in different ratios to obtain CO concentrations of 50–800 μM by using a gas mixer (High-Tech System; Bronkhorst, The Netherlands). Saturated sodium dithionite solution (10 μl) was added and the protein was injected to a final concentration of $\sim 4 \mu\text{M}$. Formation of the CO-ferrous Hb complex was verified by UV/visible absorption spectroscopy. Recombination of the photo-dissociated CO-ligand was monitored at 417 nm.

Stopped Flow Experiments

Stopped flow measurements were performed in 100 mM degassed potassium phosphate buffer (pH 7.0) and 1 mM EDTA at 20°C , by using a thermostated stopped flow apparatus (Applied Photophysics; Salisbury, UK). Sodium dithionite was added to both the protein solution and the CO solutions to a final concentration of 10 mM. Measurements were carried out during 2 s at 414 nm with 4 μM of protein solution that was mixed with different CO concentrations. Analysis was performed by using Origin software.

Nitric Oxide Dioxygenase Activities

NOD activities were assayed by following the disappearance of NO (Igamberdiev et al., 2006) with a selective electrode (ISO-NOP) coupled to a free radical analyzer (TBR4100), both from World Precision Instruments (Sarasota, FL, USA). The proteins were converted to the oxyferrous form by reduction with a trace of dithionite and rapid oxygenation through NAP-5 mini-columns (GE Healthcare Life Sciences). NOD activities were measured with diethylamine NONOate (DEA) and *S*-nitrosoglutathione (GSNO) as NO donors. DEA was purchased from Sigma and was freshly prepared for each assay. GSNO was synthesized by mixing 1 mM of acidified nitrite and 1 mM glutathione; the solution was rapidly neutralized and GSNO was quantified, aliquoted, and stored at -80°C protected from light (Smagghe et al., 2008). Concentrations of DEA and GSNO were standardized just prior to the assays by using extinction coefficients of $8 \text{ mM}^{-1} \text{ cm}^{-1}$ at 250 nm and $0.85 \text{ mM}^{-1} \text{ cm}^{-1}$ at 335 nm, respectively.

For the assay, DEA (20 μM) or GSNO (1 mM) was added to a final volume of $\sim 4 \text{ ml}$ of 50 mM potassium phosphate buffer (pH 7.5) containing 50 μM diethylenetriaminepentaacetic acid. The solution was gently stirred at 24°C until NO concentration became stable ($\sim 6 \mu\text{M}$ with DEA and $\sim 2 \mu\text{M}$ with GSNO after $\sim 4 \text{ min}$). The oxyferrous Hb (1 μM ; 30–60 μl) was added, while stirring, so that the final volume of the reaction mixture was exactly 4 ml, and the decrease in NO concentration was measured. The time between the preparation of oxyferrous Hbs and the assays of NOD activity was always $< 5 \text{ min}$. The corresponding ferric globins lacked NOD activity and were employed as negative controls. The NO electrode was calibrated for each set of measurements by following the manufacturer's instructions.

RESULTS

Purification of nsHbs and Identification of Disulfide Bond

Recombinant LjGlbs, and the mutant derivatives LjGlb1-1 C8S, LjGlb1-1 C78S, and LjGlb1-2 C79S, were highly purified and the protein preparations usually exhibited Soret/ A_{280} ratios > 2.8 . Unfortunately, we were unable to produce the LjGlb2 C65S at enough yield for kinetic and structural studies because of the instability of the protein. We found that LjGlb1-2 and LjGlb2 are monomeric proteins. However, LjGlb1-1 was present

both as a monomer and dimer when purified in the presence of 200 mM NaCl, whereas only the monomer was found when the salt was omitted during purification. The homodimer was formed by a disulfide bond, as revealed by FPLC and mass spectrometry analysis in the absence and presence of dithiothreitol (Supplementary Figure S2). The disulfide bridge involves Cys8 because the LjGlb1-1 C78S mutant is still able to form a dimer that disappears upon addition of dithiothreitol (data not shown). Interestingly, barley Hb1 is a homodimer having a disulfide bond through its unique residue and hence the protein is stable without salt (Duff et al., 1997; Bykova et al., 2006), whereas rice Hb1 is also a homodimer but does not appear to form disulfide bridges (Goodman and Hargrove, 2001). We found that the dimer of LjGlb1-1 precipitated if salt was omitted during purification and the spectroscopic studies of the wild-type LjGlb1-1 and its mutated forms were therefore performed in buffer supplemented with 200 mM NaCl.

Resonance Raman Spectroscopy

Earlier work has shown that RR spectroscopy is most useful to identify different oxidation and ligation states of the globins and to study in detail the stabilization of heme ligands by the amino acid residues in the heme pocket (Hu et al., 1996). Accordingly, we have used RR to compare the heme environments in the wild-type and mutant LjGlbs.

The RR spectra were obtained in the high-frequency region (1300–1650 cm^{-1}), which contains marker bands for the oxidation, coordination, and spin state of the heme iron, as well as in the low-frequency region (200–700 cm^{-1}), which contains several in-plane and out-of-plane vibrational modes of the heme. **Figure 1** shows the RR spectra of all the wild-type proteins and their mutated versions in their deoxyferrous form. All RR spectra show marker bands ν_4 at 1361–1363 cm^{-1} and ν_3 at 1493–1496 cm^{-1} , which are characteristic for a hexacoordinate low-spin (6cLS) ferrous form. Additionally, a second ν_3 band is seen at 1467–1475 cm^{-1} , which indicates the presence of a pentacoordinate high-spin (5cHS) ferrous form. This band has a substantially larger intensity for LjGlb2 relative to LjGlb1-1 and LjGlb1-2. Although the intensities of the two ν_3 marker bands seem to be similar for LjGlb1-1 and LjGlb1-2, the population of the pentacoordinate (5c) ferrous heme is nevertheless smaller than that of the hexacoordinate (6c) ferrous heme. Indeed, the

intrinsic intensity of the ν_3 marker band of 5c heme is much higher than that of 6c heme (Das et al., 1999), which makes it difficult to get accurate relative populations out of the RR spectra. This is confirmed by the optical absorption spectra that were shown in earlier work (Sainz et al., 2013), which indicated that in the ferrous forms of LjGlb1-1 and LjGlb1-2, a mixture between 5cHS and 6cLS is observed, with the latter being the dominant species. On the contrary, in the ferrous form of LjGlb2, the 5cHS species prevails over the 6cLS species, in line with earlier absorption measurements (Sainz et al., 2013). Taken together, the RR and UV/visible spectra reveal that, for LjGlb1-1 and LjGlb1-2, a 5cHS form is present as a minor fraction, whereas this constitutes the most abundant fraction in LjGlb2. The equilibrium between the 5cHS and 6cLS species, evidenced by the spectroscopic data, is confirmed by the distal His binding constants reported in **Table 1**. The high-frequency region of the RR spectra of the deoxyferrous form of both LjGlb1-1 and LjGlb1-2 is similar to that found for barley Hb1 (Das et al., 1999), tomato Hb1 (Ioanitescu et al., 2005), and AtGlb2 (Bruno et al., 2007a), which all have a bis-histidyl coordination of the heme. This is in contrast with LjGlb2, which has a mixed 5cHS–6cLS state in the ferrous form and is more similar to AtGlb1 (Bruno et al., 2007a). Mutation of Cys to Ser alters the relative intensity ratio of the ν_3 marker bands for LjGlb1-1 (**Figures 1A–C**). More specifically, the relative fraction of 5cHS increases for LjGlb1-1 C8S and decreases for LjGlb1-1 C78S when compared to the wild-type protein. Overlay of the RR spectra of LjGlb1-2 and LjGlb1-2 C79S shows a small decrease in the 5cHS contribution in the mutant (**Figures 1D,E**).

The low-frequency region (200–730 cm^{-1}) of the RR spectra contains a number of bending modes from the vinyl and propionate substituents of the heme group. The propionate bending mode [$\delta(\text{C}_\beta - \text{C}_\alpha - \text{C}_\alpha)$] appears at 382–384 cm^{-1} . For comparison, this mode is found at 380 cm^{-1} in ferrous barley Hb (Das et al., 1999) and tomato Hb (Ioanitescu et al., 2005). It is possible to use this mode to quantify the strength of the interaction between the heme propionate groups and nearby amino acid residues; a higher Raman shift for the propionate bending mode indicates a stronger interaction (Cerdeña-Colon et al., 1998). This interaction seems to be similar for all studied LjGlbs. When compared with ferrous barley and tomato Hbs, the interaction seems stronger for the LjGlbs. The vinyl bending

TABLE 1 | Rate constants of LjGlbs and their mutated derivatives.

Globin	$k_{\text{on,CO}}^a$ ($\mu\text{M}^{-1} \text{s}^{-1}$)	$k_{\text{on,H}}^b$ (s^{-1})	$k_{\text{off,H}}^{a,b}$ (s^{-1})	K_H^c	F_H^d	F_{gem}^e	k_{gem}^f ($\times 10^7 \text{s}^{-1}$)
LjGlb1-1	1.2	67	27	2.48	0.71	0.09	9.2
LjGlb1-1 C8S	1.6	111	57	1.95	0.66	0.04	6.8
LjGlb1-1 C78S	1.7	288	5.1	56.8	0.98	0.13	6.2
LjGlb1-2	29	1484	88	16.9	0.94	0.27	7.6
LjGlb1-2 C79S	67	3501/1051	84/5.8	41.7/181	0.98/0.99	0.46	8.8
LjGlb2 ^g	3.79 (70%) 10.71 (30%)	0.033/4.4	0.15/20	0.22	0.18	0.15	5.7

^aDetermined by LFP. ^bDetermined by stopped flow. ^cEquilibrium constant for His binding $K_H = k_{\text{on,H}}/k_{\text{off,H}}$. ^d $F_H = K_H/(1 + K_H)$, fraction of bis-histidyl 6c species in ferrous proteins at equilibrium. ^e F_{gem} , fractional amplitude of geminate rebinding. ^f k_{gem} , rate constant for geminate rebinding. ^gFor LjGlb2, the hexacoordinated fraction F_H is calculated from the absorption spectrum and is consistent with the amplitude of binding to the 5c species in stopped flow. Because $K_H = F_H/(1 - F_H)$, from $F_H = 0.18$ we obtain $K_H = 0.22$. There are two $k_{\text{off,H}}$ values (0.15 and 20 s^{-1}). Assuming the same equilibrium constant, we can calculate two values for $k_{\text{on,H}} = K_H \cdot k_{\text{off,H}}$ (0.033 and 4.4 s^{-1} , respectively).

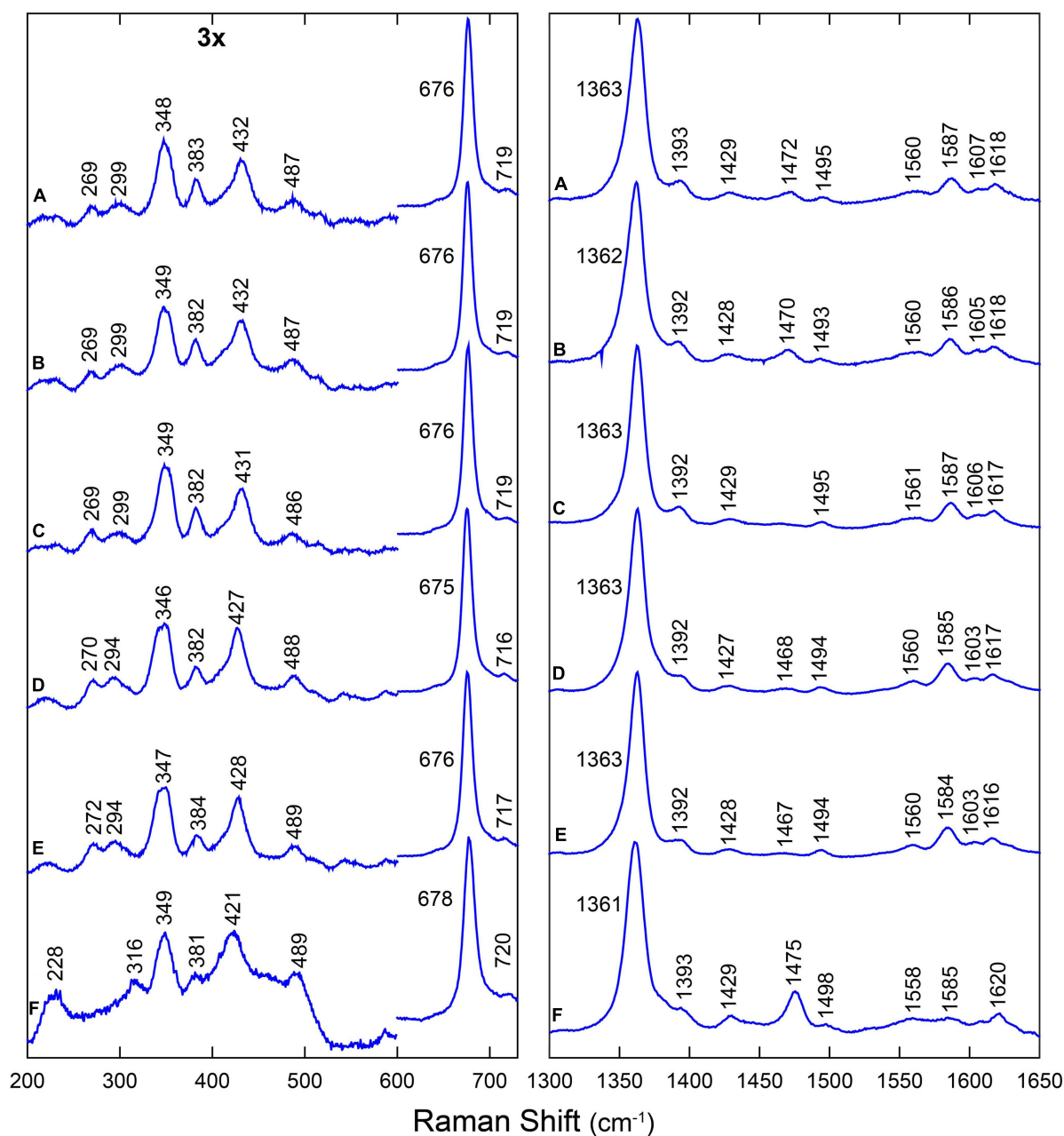


FIGURE 1 | RR spectra of LjGlb1s in the deoxyferrous state. The spectra correspond to (A) LjGlb1-1, (B) LjGlb1-1 C8S, (C) LjGlb1-1 C78S, (D) LjGlb1-2, (E) LjGlb1-2 C79S, and (F) LjGlb2. All the spectra were recorded with a laser power of 12 mW.

modes [$\delta(C_\beta - C_\alpha - C_\beta)$] are seen as a single line at 428–432 cm^{-1} (LjGlb1-1 and LjGlb1-2) and 422 cm^{-1} (LjGlb2). This indicates a stronger interaction of the vinyl group with surrounding amino acid residues for both class 1 Hbs. In barley Hb the vinyl bending modes were found at 425 cm^{-1} . The overlap of both bending modes of the vinyl groups indicates a relaxed heme configuration. The γ_7 pyrrole bending mode, which is associated with a heme out-of-plane distortion, is only visible for LjGlb2 at 316 cm^{-1} (Figure 1F), in line with the relative

high fraction of 5cHS heme in this protein. Both LjGlb1-1 and LjGlb1-2 lack out-of-plane modes γ_6 , γ_7 , γ_{12} , and γ_{21} , which is typical for a bis-histidyl coordination. This indicates, together with the overlap of both bending modes of the vinyl groups, that for the class 1 nsHbs, the heme is in a relaxed state with the heme iron almost completely in the porphyrin plane. Finally, the $\nu_{\text{Fe-His}}$ stretching mode (228 cm^{-1}) is only visible for LjGlb2 in the deoxyferrous state. This is in agreement with the presence of the 5cHS form, since the Fe-His stretching mode is generally

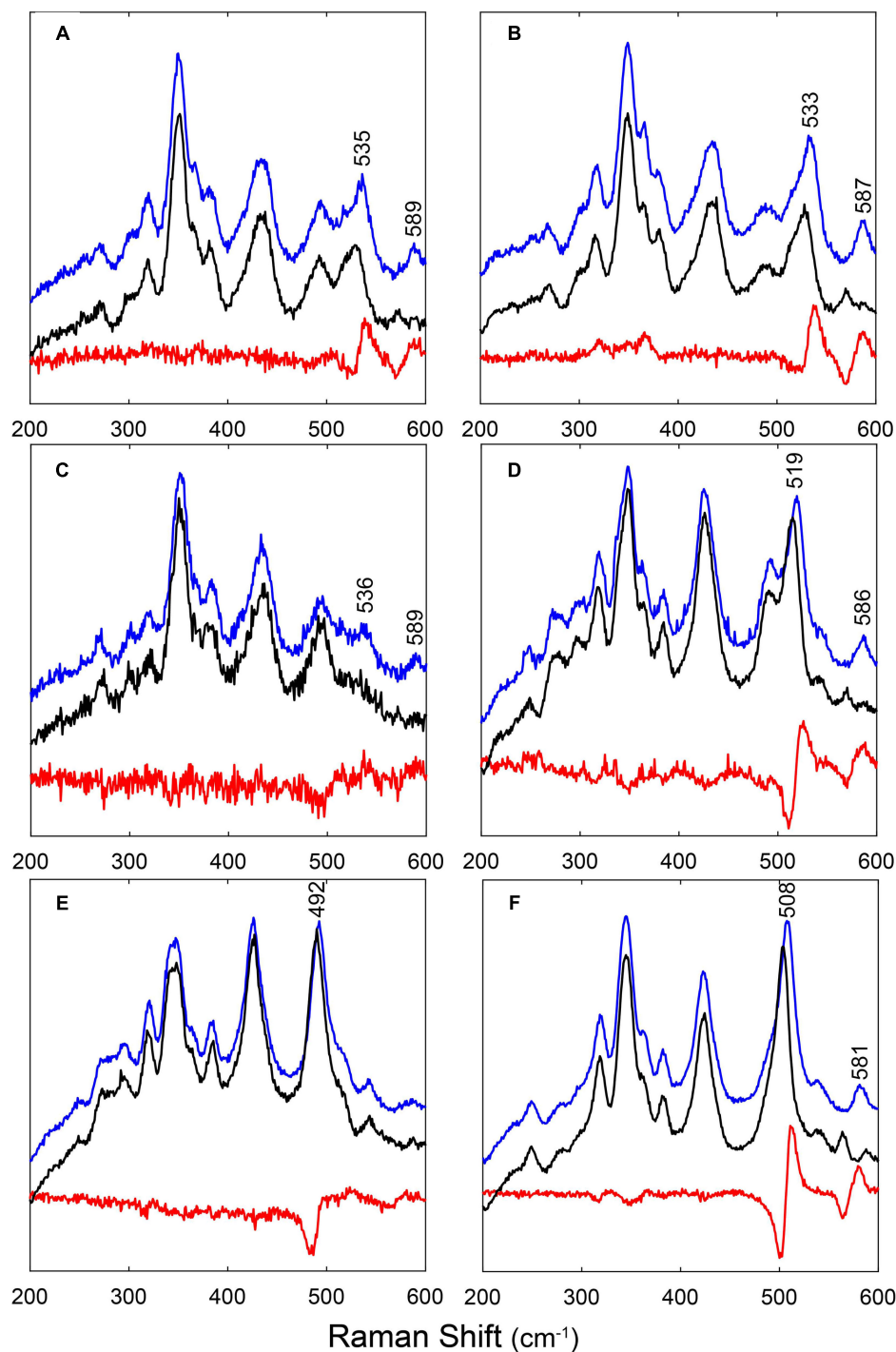
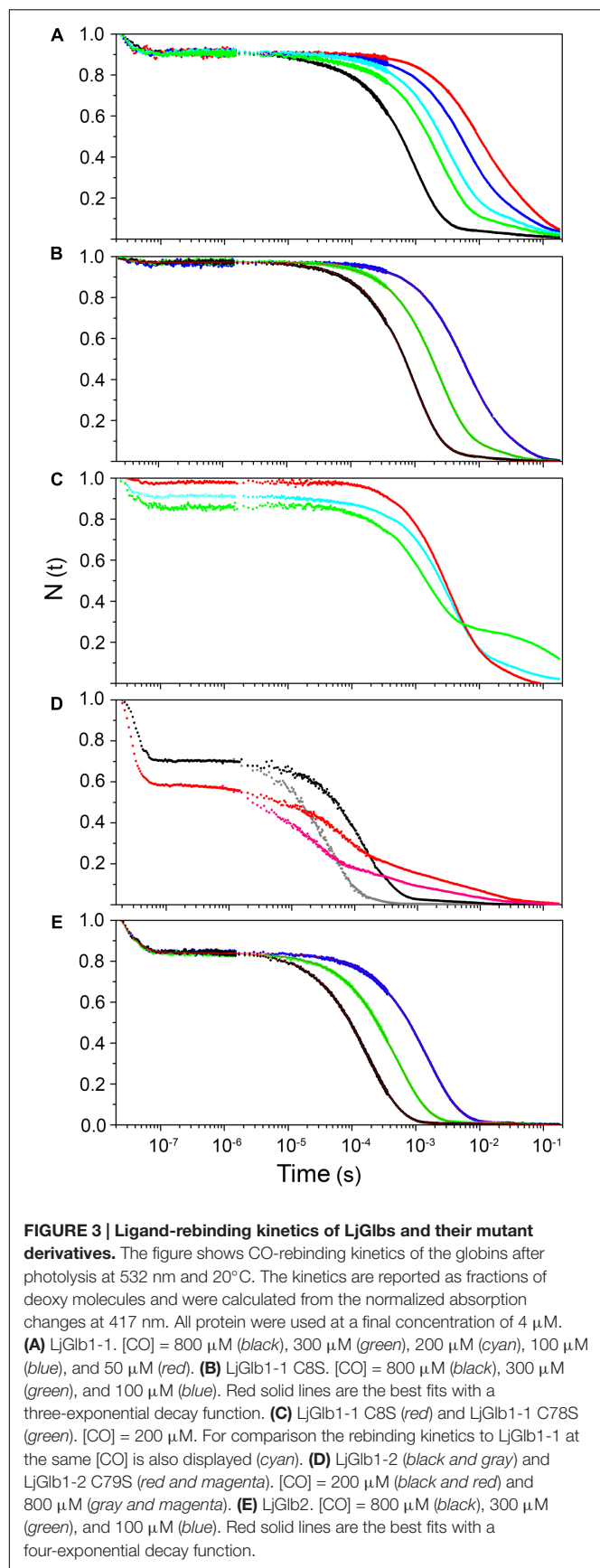


FIGURE 2 | RR spectra of LjGlbs and of their mutant derivatives in the ferrous CO-ligated form. The figure shows the spectra of (A) LjGlb1-1, (B) LjGlb1-1 C8S, (C) LjGlb1-1 C78S, (D) LjGlb1-2, (E) LjGlb1-2 C79S, and (F) LjGlb2. The upper (blue), middle (black), and lower (red) spectra correspond to the ^{12}CO form, the ^{13}CO form, and the difference spectra, respectively. All spectra were recorded with a laser power of 1 mW.

not observed for bis-histidyl coordinated globins. The value of $\nu_{\text{Fe-His}}$ of LjGlb2 is somewhat higher than that observed for barley Hb (219 cm^{-1}) but still typical for globins (Das et al., 1999).

Figure 2 shows the RR spectra of the ferrous $^{12/13}\text{CO}$ -ligated forms of the globins and their mutated forms. In addition, Supplementary Figure S3 includes a comparison of the RR spectra of the ferrous CO-ligated forms of all proteins recorded with low



(1 mW) and high (35–165 mW) laser power. Upon increasing the laser power, partial photolysis occurs which is apparent from the shift of the ν_4 frequency from the CO-ligated ($\sim 1375\text{ cm}^{-1}$) to the ferrous ($\sim 1362\text{ cm}^{-1}$) state. Because of this photolysis, a general decrease in intensity of the Fe-CO stretching modes ($\nu_{\text{Fe-CO}}$) is expected. We clearly see a spectral change in the $450\text{--}600\text{ cm}^{-1}$ region, where the Fe-CO stretching modes occur. Further assignment of these bands is corroborated by comparing the RR spectra of the ^{12}CO - and ^{13}CO -ligated globins (Figure 2). The use of ^{13}CO induces a downward shift in the $\nu_{\text{Fe-CO}}$ modes of $\sim 3\text{ cm}^{-1}$, whereas the Fe-CO bending mode ($\delta_{\text{Fe-CO}}$) mode shifts from $582\text{--}589\text{ cm}^{-1}$ to $564\text{--}569\text{ cm}^{-1}$. This can be better seen in the difference spectrum ($^{12}\text{CO}\text{--}^{13}\text{CO}$) (Figure 2; red spectra).

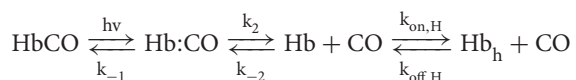
The $\nu_{\text{Fe-CO}}$ modes are sensitive to interactions of the CO ligand with nearby amino acid residues. Whereas Fe-CO stretching modes around $490\text{--}495\text{ cm}^{-1}$ indicate an open heme pocket in which the CO interacts only weakly with the surrounding amino acids, higher $\nu_{\text{Fe-CO}}$ modes are due to a closed heme pocket in which a positively charged amino acid residue is stabilizing the CO group (Spiro and Wasbotten, 2005; Bruno et al., 2007a). In general, the higher the mode, the stronger the CO ligand will be hydrogen bonded and interact with the positively charged residue. For LjGlb1-1 and its C8S and C78S mutants, the $\nu_{\text{Fe-CO}}$ modes were found at $\sim 533\text{--}536\text{ cm}^{-1}$, with $\delta_{\text{Fe-CO}}$ at $587\text{--}589\text{ cm}^{-1}$ (Figures 2A–C and Supplementary Figure S3a–c). This indicates a very strong interaction of the CO ligand, which is not affected by the mutations. In contrast, the $\nu_{\text{Fe-CO}}$ mode is at 519 cm^{-1} for wild-type LjGlb1-2, but at 492 cm^{-1} for its C79S mutant (Figures 2D,E and Supplementary Figure S3d,e). In this case, the mutation induces a switch from a closed to an open heme pocket. Consistent with this, the Fe-CO bending mode is clearly visible for the closed configuration ($\sim 586\text{ cm}^{-1}$) of the wild-type protein, but it is hardly observed for the open heme pocket of the mutant (Figure 2E). Finally, the $\nu_{\text{Fe-CO}}$ mode of LjGlb2 at 508 cm^{-1} (Figure 2F and Supplementary Figure S3f) is similar to that observed for CO-ligated vertebrate Mbs, where the CO stabilization occurs through the His(E7) residue. The interaction of bound CO with the surrounding amino acid residues is thus weaker in class 2 than in class 1 Hbs, in line with the occurrence of a dominant fraction of 5cHS form in the deoxyferrous state of LjGlb2 (Figure 1F).

Ligand Binding Kinetics

As evidenced from the RR experiments (Supplementary Figure S3), the Fe-CO bond in hemeproteins is photolabile. LFP exploits this property to photodissociate the ligand with a short (nanosecond) laser pulse and monitor rebinding through the concomitant absorption changes. Photodissociated ligands can either be rebound by the heme from temporary docking sites within the protein matrix (geminate rebinding) or migrate to the solvent and be rebound at later times (bimolecular rebinding) (Abbruzzetti et al., 2006). The CO rebinding kinetics of LjGlb1-1, LjGlb1-2, and LjGlb2 were examined by LFP (Figure 3). For LjGlb1-1, a geminate phase was observed in the nanosecond range, which accounts for $\sim 10\%$ of the rebinding and is well described by a single exponential relaxation (Figure 3A). On

longer time scales, the progress curve for LjGlb1-1 is dominated by a large microsecond to millisecond phase with two easily recognizable kinetic steps (**Figure 3A**). The faster of these two steps has a clear bimolecular nature as demonstrated by the response of the kinetics to ligand concentration. When [CO] is decreased, the apparent rate of the faster step becomes lower, as expected for a diffusion-mediated bimolecular reaction. Accordingly, this step is identified as the reaction between CO and the LjGlb1-1 5c species. On the other hand, the amplitude of the slower step becomes higher when [CO] is decreased, whereas the apparent decay rate is unaffected. This is consistent with the transient formation of the bis-histidyl 6c species, which is observed when the distal His(E7) is coordinated to the sixth coordination site at the heme Fe, made available by the photolysis of the parent CO adduct. At lower [CO], the slower bimolecular rebinding to the 5c heme allows for a more efficient relaxation toward the bis-histidyl 6c species, thus resulting in a larger accumulation of this intermediate. Eventually, His(E7) will be displaced by CO, with a rate which is independent of [CO] and coincides with the His dissociation rate.

The overall kinetics is well described by the sum of three exponential decays, corresponding to the kinetic phases described above, and which are, in order of increasing lifetime: geminate rebinding, bimolecular binding to 5c species, and decay of bis-histidyl 6c species.



The above scheme summarizes the relevant kinetic steps and the corresponding microscopic rate constants. After photodissociation with a photon of energy $h\nu$, the ligand can be rebound geminately from positions located within the protein matrix (Hb:CO) with rate k_{-1} , or escape to the solvent (Hb_p + CO) with rate k_2 . Ligands can be rebound from the solvent with rate k_{-2} . On the same time scale, the distal His can coordinate the Fe atom by forming a bis-histidyl hexacoordinate species Hb_h with rate $k_{\text{on,H}}$. This species eventually decays with rate $k_{\text{off,H}}$.

The bimolecular CO rebinding rate to the 5c protein, $k_{\text{on,CO}}$, is related to microscopic rate constants according to this equation:

$$k_{\text{on,CO}} = k_{-2} \frac{k_{-1}}{k_{-1} + k_2} \quad (1)$$

The fitting, performed on a set of rebinding traces comprising five different CO concentrations, afforded to calculate apparent rates for each process. As expected, no difference in amplitude or apparent rate was observed for the geminate phase. On the contrary, the apparent rate for the bimolecular reaction between CO and the 5c heme increases linearly with [CO], and from the slope we could estimate the $k_{\text{on,CO}}$ for the process (**Table 1**). The slowest process affords an estimate for the $k_{\text{off,H}}$ rate. The two Cys residues of LjGlb1-1 appear to play a role in the overall structure and dynamics of the protein, with functional consequences on the heme ligand-binding kinetics (**Figures 3B,C**). Moreover, the C8S and C78S

mutations induce very different effects on the overall CO rebinding kinetics. The C8S mutant is characterized by a smaller geminate amplitude (4%) with a minor decrease in the corresponding rate. The $k_{\text{on,CO}}$ and the $k_{\text{off,H}}$ are not substantially affected, whereas the extent of 6c species becomes a bit smaller, indicating a lower binding rate for His(E7). On the contrary, the C78S mutation leads to larger geminate recombination (13%), faster $k_{\text{on,CO}}$, and slower $k_{\text{off,H}}$. The larger accumulation of bis-histidyl species also indicates a larger $k_{\text{on,H}}$. The changes in geminate recombination of the C8S and C78S mutants suggest that the egression pathway is affected by mutations at these two residues, albeit in the opposite direction.

The CO rebinding kinetics of LjGlb1-2 (**Figure 3D**) shares some similarity to that of LjGlb1-1, with a nanosecond geminate recombination followed by a biphasic kinetics. Thus, it is possible to recognize a bimolecular step corresponding to CO rebinding to a 5c species and a slower step associated with the decay of the bis-histidyl species. However, amplitudes and rates are dramatically different from the ones determined for LjGlb1-1. Geminate rebinding to LjGlb1-2 is a prominent process accounting for ~30% of the kinetics, with an apparent rate which is not substantially different from that determined for LjGlb1-1. A minor contribution from a second transient in the bimolecular phase is detected in all traces for LjGlb1-2. Given the small amplitude and the irregular trend, it is ascribed to an impurity and neglected in the current analysis. The apparent rates for the bimolecular phase are clearly higher for this globin, as can be easily appreciated by visual inspection of the traces corresponding to the same [CO] (compare **Figures 3A,D**). The decay of the bis-histidyl species also appears faster than for LjGlb1-1. The kinetics are well reproduced by a sum of three exponential decays at all investigated values of [CO]. The fitting parameters reported in **Table 1** reflect the above qualitative description. The amplitude of geminate rebinding for LjGlb1-2 increases to ~30%, indicating that for photodissociated ligands it is more difficult to reach the solvent than in the case of LjGlb1-1. A much higher value for $k_{\text{on,CO}}$ is observed, in keeping with the higher geminate rebinding, and the decay of the bis-histidyl species is also a faster process. The C79S mutation has profound consequences on the rebinding kinetics (**Figure 3D**), with higher geminate rebinding, faster bimolecular rebinding, and accumulation of a much higher population of the bis-histidyl 6c species. Accordingly, the marker band of 5cHS in the RR spectrum decreased for this mutant (**Figure 1E**). **Table 1** shows that the amplitude of geminate rebinding increases to 46%, and that $k_{\text{on,CO}}$ undergoes a twofold increase. The decay rate of the bis-histidyl 6c species is not substantially affected. Because this species is accumulated in higher yield, it is expected that the on-rate for the process will be higher.

The progress curve for CO rebinding to LjGlb2 shows the general kinetic pattern already highlighted for the class 1 nsHbs previously discussed, and is well described by a sum of exponential decays (**Figure 3E**). However, unlike the other globins described in this work, two exponential decays are needed to properly account for the bimolecular phase, a fact that may arise from two conformations coexisting in equilibrium. Their

$k_{\text{on,CO}}$ values differ by about twofold (Table 1). Although the long time tail of the rebinding kinetics is barely appreciable in the plot of Figure 3E, this kinetic phase has the typical features of the decay of bis-histidyl 6c species. The small amount of this intermediate is indicative of a very low, but not negligible, binding rate for the distal His.

Stopped flow, rapid mixing experiments were conducted with all nsHbs to determine the $k_{\text{on,H}}$ and $k_{\text{off,H}}$ rates. When deoxyferrous Hb solutions are mixed with a solution equilibrated with CO, the exogenous ligand is bound by the protein in a bimolecular reaction. In the 6c proteins like nsHbs, however, the endogenous ligand His(E7) must first dissociate from the heme so that the diatomic gas is able to bind to the heme. At high enough [CO], this step becomes rate limiting. For Hbs that are only partly hexacoordinated (in which a fraction of 5c species is present at equilibrium), binding of CO in the rapid mixing experiments is a biexponential process described by the following equation:

$$\Delta A_{\text{obs}} = -A_T \left(F_P e^{-k_{\text{on,CO}}[\text{CO}]t} + F_H e^{-k_{\text{obs}}[\text{CO}]t} \right) \quad (2)$$

In this equation, ΔA_{obs} is the observed parameter for binding; $k_{\text{on,CO}}$ is the bimolecular rate constant for CO binding to the 5c species; k_{obs} is the observed rate constant for binding following mixing; F_P and F_H are the fractions of protein in the 5c and 6c states; and A_T is the total change in absorbance expected for the reaction, determined independently from ligand-free and ligand-bound absorbance spectra (Smaghe et al., 2006). The equation that describes the apparent rate k_{obs} for these kinetics is as follows (Trent et al., 2001):

$$k_{\text{obs}} = \frac{k_{\text{off,H}} k_{\text{on,CO}} [\text{CO}]}{k_{\text{on,H}} + k_{\text{off,H}} + k_{\text{on,CO}} [\text{CO}]} \quad (3)$$

Figure 4A shows the progress curves for CO binding to LjGlb1-1 at several values of [CO], along with fits using double exponential relaxations. Figure 4B reports the [CO] dependence of the apparent rate constant associated with CO binding to the 6c species. Like other 6c globins, a typical trend is observed for the slow kinetic phase, where the apparent rate constant reaches a saturating value at high [CO] (Smaghe et al., 2006). This limiting value corresponds to the $k_{\text{off,H}}$ rate. For all the proteins, the trend of the rates k_{obs} with [CO] is well described by Eq. 2, where $k_{\text{on,H}}$ and $k_{\text{off,H}}$ are held as free parameters, whereas the value of $k_{\text{on,CO}}$, determined from flash photolysis, is held as a fixed parameter. The retrieved parameters for LjGlb1-1 and the other proteins are reported in Table 1.

Figure 5A compares the expected values for k_{obs} using the model gas CO for several nsHbs. It is quite clear that, due to the combination of rates $k_{\text{on,CO}}$, $k_{\text{on,H}}$, and $k_{\text{off,H}}$, LjGlb1-2 binds CO with higher rate than LjGlb1-1 in these conditions. The k_{obs} values for LjGlb1-2 and its C79S mutant (Figure 5B) are quite similar. On the contrary, a fivefold decrease in the rate is observed for the C78S mutant of LjGlb1-1, which suggests a critical role of this residue in determining the rate constants relevant for k_{obs} . In contrast, the effect of the C8S mutation appears to be negligible.

The values of the equilibrium constants (K_H) for the binding of the distal His(E7) to the hemes of the three LjGlbs are shown in Table 1. The K_H of LjGlb1-1 demonstrates that a fraction of 5c species is present at equilibrium. This is in keeping with the presence of a small intensity, second ν_3 band observed at 1467–1475 cm^{-1} in the RR spectrum (Figure 1A), indicating the presence of a 5cHS ferrous form. The LjGlb1-1 C8S mutation slightly shifts the equilibrium toward the 5cHS species. The opposite effect is observed for the LjGlb1-1 C78S mutant. The weak second ν_3 band observed at 1467–1475 cm^{-1} appears to behave consistently (Figures 1B,C). The K_H of LjGlb1-2 clearly indicates that the unliganded ferrous form of the protein is mostly present as the bis-histidyl 6c species, as anticipated by the absorption spectra (Sainz et al., 2013) and the RR spectra (Figure 1). The LjGlb1-2 C79S mutation results in heterogeneous kinetics in the time range where the bis-histidyl species is formed and decays, where two exponential decays are needed to account for the measured time course of CO binding. Plotting the two rate constants as a function of [CO] and fitting their trend using Eq. 2 allows to retrieve the $k_{\text{on,H}}$ and $k_{\text{off,H}}$ values reported in Table 1. For both conformations, these rates result in stronger hexacoordination than observed for the wild-type protein. The reason for the presence of the two species is as yet unclear. Finally, the K_H of LjGlb2 is quite low, indicating a substantially lower stability of the bis-histidyl 6c species. Consistently, RR spectra of LjGlb2 (Figure 1) show a remarkably high population of 5cHS species, and the absorption spectra clearly show a mixture of 5c and 6c species (Sainz et al., 2013).

Nitric Oxide Dioxygenase Activities

Because both class 1 and class 2 nsHbs are able to scavenge NO *in vitro* and *in vivo* (Perazzolli et al., 2004; Igamberdiev et al., 2006; Hebelstrup and Jensen, 2008; Smaghe et al., 2008), we measured NOD activities of LjGlbs, as well as of some of their mutated forms, to identify possible differences among the proteins and to determine whether Cys residues play a role in NO scavenging activity. To this end, we used a well-known artificial NO donor (DEA) and a physiological NO donor (GSNO). At the concentrations employed, both compounds released NO linearly for ~4 min, at which time NO concentration stabilized. The oxyferrous Hbs were then added and the initial rate of NO consumption was measured and expressed on the basis of heme protein concentration (Figure 6). The decrease in NO concentration was due to NOD activity mediated by the hemes and was unrelated to scavenging by Cys residues because the respective ferric Hbs had no activity. We found that wild-type LjGlb1-1, LjGlb1-2, and LjGlb2 exhibited similar NOD activities regardless of the NO donor. Also, there were no major differences of NOD activity between the wild-type and the mutated proteins, although the LjGlb1-1 mutants displayed higher NOD activity than the wild-type LjGlb1-1. The cause for this minor, yet significant, increase is uncertain because it did not occur in LjGlb1-2 C79S (Figure 6), which again indicates that the Cys residues are not involved in NO scavenging.

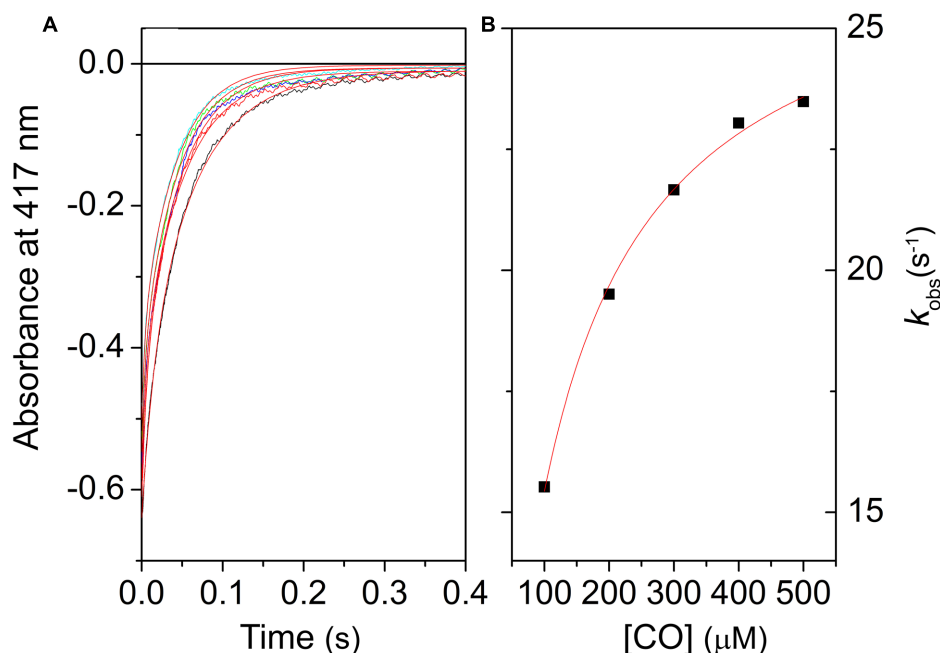


FIGURE 4 | Kinetics of CO binding to LjGlb1-1 observed after stopped-flow rapid mixing. (A) Absorbance changes at 414 nm following rapid mixing of deoxyferrous LjGlb1-1 in CO-equilibrated buffer. Final [CO] after mixing was 100 μM (black), 200 μM (red), 300 μM (green), 400 μM (blue), and 500 μM (cyan). Red solid curves are the best fits with double exponential relaxations. **(B)** Observed rate constants for the slow kinetic phase of LjGlb1-1 as a function of [CO]. The rates of the fast kinetic phase exhibit a linear trend with slope $k_{\text{on,CO}}$ (not shown in the figure). The red solid line is the best fit of the data with Eq. 2.

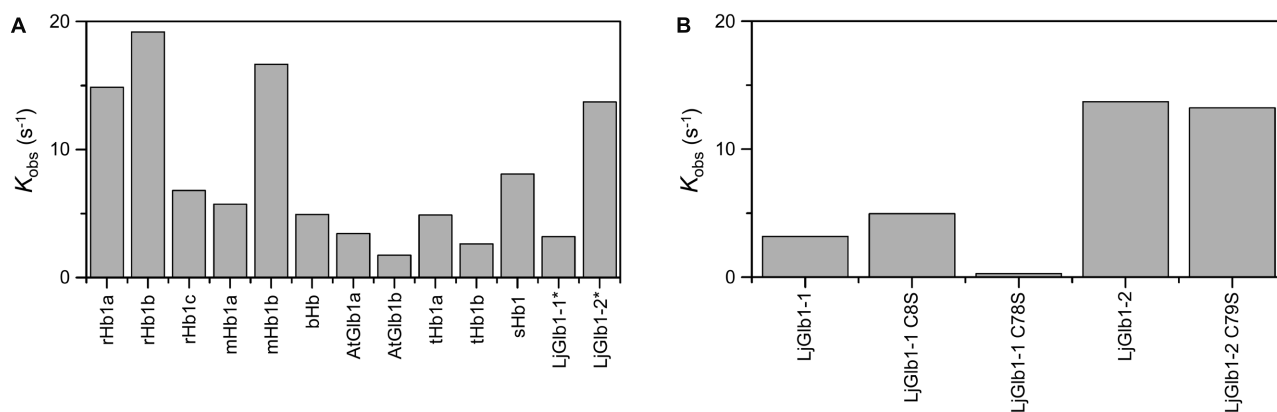


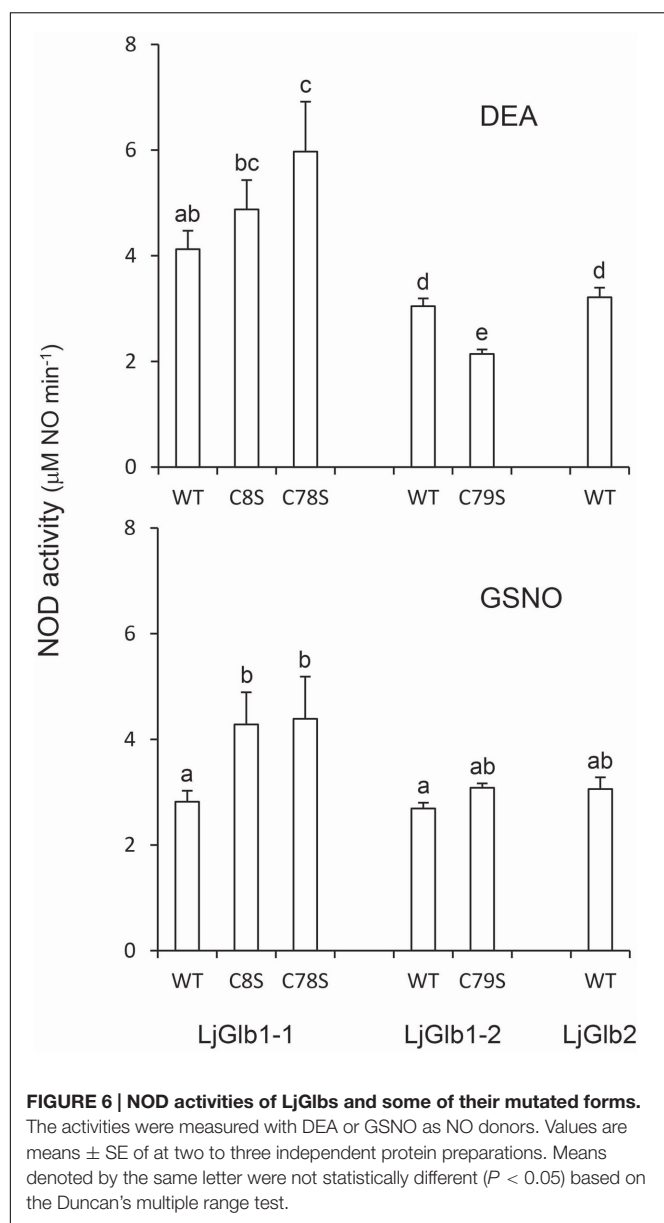
FIGURE 5 | Apparent binding rates of CO to several plant nsHbs. (A) Comparison of apparent rates (k_{obs}) determined from the values of $k_{\text{on,CO}}$, $k_{\text{on,H}}$, and $k_{\text{off,H}}$ reported in the literature. [CO] = 10 μM. Abbreviations and references: rice Hb1 (rHb1a; Arredondo-Peter et al., 1997; Bisht et al., 2011), rice Hb1 (rHb1b; Smaghe et al., 2006), maize Hb1 (mHb1a; Smaghe et al., 2006), maize Hb1b (mHb1b; Smaghe et al., 2006), barley Hb1 (bHb1; Duff et al., 1997), *Arabidopsis thaliana* Glb1 (AtGlb1; Bruno et al., 2007a; Smaghe et al., 2009), tomato Hb1 (tHb1; Ioanitecu et al., 2005; Smaghe et al., 2009), soybean Hb1 (sHb1; Smaghe et al., 2009). Values of LjGlb1-1 and LjGlb1-2 (marked with asterisks) were obtained for this work. For rice and maize, values of two different class 1 Hbs, termed “a” and “b,” are provided. **(B)** Comparison of k_{obs} values of LjGlb1-1, LjGlb1-2, and their mutant derivatives.

DISCUSSION

In this work, the electronic and ligand binding properties of the heme environments of the three nsHbs of *L. japonicus* were examined by combining several spectroscopies and assaying the NOD activities of the proteins. The comparisons between LjGlb1-1 and LjGlb1-2, as well as between class 1 and class 2

nsHbs, were facilitated by using mutated proteins, which enabled us to determine the effect of Cys residues on protein stability and ligand affinity. This type of studies on plant nsHbs is scarce, yet important to understand protein function.

Although the general behavior of the two class 1 nsHbs is similar, the details of their distal pocket and the ligand rebinding kinetics show significant differences. For LjGlb1-1, the geminate



phase of ligand binding is comparable to that of other class 1 nsHbs. Thus, the CO rebinding to rice Hb1 is characterized by $\sim 10\%$ geminate rebinding that occurs with a nearly mono-exponential kinetics. Similarly, the geminate rebinding to AtGlb1 is a process with a comparable amplitude, but a slightly more complex kinetics that is well described by a bi-exponential relaxation (Abbruzzetti et al., 2006). In AtGlb1 the bi-exponential nature of the kinetics was interpreted as a result of the migration of the photodissociated ligand to nearby cavities, from which the ligand is rebound at later times with different rates (Abbruzzetti et al., 2007; Bruno et al., 2007a). Both the equilibrium binding constants and the binding rates to hemeproteins are profoundly influenced by structural properties of the active site, including the presence of temporary docking sites within the protein matrix, and tunnels connecting the interior of the protein with the

solvent. The strong interaction of the CO ligand with the distal His residue observed for LjGlb1-1 and its C8S and C78S mutants, for which the $\nu_{\text{Fe-CO}}$ modes were found at $\sim 533\text{--}536\text{ cm}^{-1}$ and $\delta_{\text{Fe-CO}}$ at $587\text{--}589\text{ cm}^{-1}$ (Figure 2 and Supplementary Figure S3), may be taken as the main reason for the high affinity of this protein for diatomic gases, because this interaction is expected to decrease substantially the ligand dissociation rate constant. Notably, the same values (535 cm^{-1}) were found for barley Hb1 (Das et al., 1999) and AtGlb1 (Bruno et al., 2007a). The K_{H} value of LjGlb1-1 is consistent with those of other class 1 nsHbs (Supplementary Figure S4A). It is interesting to note that while $k_{\text{on,H}}$ and $k_{\text{off,H}}$ show a large variability across the class 1 nsHbs, these rate constants appear to be strongly correlated ($r = 0.87$; Supplementary Figure S4B). The slope provides a K_{H} of 1.6 ± 0.2 , a value that implies partial hexacoordination for the equilibrium, ligand-free deoxyferrous species. A similar average value of 1.7 ± 0.2 , estimated from the analysis of several class 1 nsHbs, was reported earlier (Smaghe et al., 2009). The reasons for the correlation may be ascribed to the dynamics of the different proteins and may involve an enthalpy-entropy compensation. The transition from the bis-histidyl hexacoordination to pentacoordination implies conformational changes of the protein. Central to this conformational change is the peculiar translation of helix E along its axis. The flexibility of the CD and EF loop regions in class 1 nsHbs allows the piston motion of the E-helix that accompanies dissociation of the distal His and subsequent ligand binding. This flexibility, along with the unfavorable interaction of Phe(B10) with the coordinated distal His, promotes reversible hexacoordination (Hoy et al., 2007).

LjGlb1-2 is clearly an outlier with respect to CO binding rate, His binding and dissociation rates, and K_{H} values, which are much higher than the typical values observed for class 1 nsHbs. The amplitude of geminate rebinding for LjGlb1-2 is much larger than the typical values reported for other class 1 nsHbs and, in fact, it is similar to the amplitude observed for AtGlb2 (Bruno et al., 2007a,b). Likewise, the $k_{\text{on,CO}}$ of LjGlb1-2 resembles the values of class 2 nsHbs rather than those of class 1 nsHbs (Supplementary Figure S5). The distal pocket of LjGlb1-2 is quite closed, with $\nu_{\text{Fe-CO}}$ mode at 519 cm^{-1} , suggestive of a strong interaction with distal pocket residues, but this interaction is less strong than for LjGlb1-1. Unlike the case of LjGlb1-1, the presence of a closed distal pocket in LjGlb1-2 does not impair geminate recombination, which occurs with a remarkably large amplitude ($\sim 30\%$), probably due to the weaker interaction. The $\nu_{\text{Fe-CO}}$ mode undergoes a dramatic change for the LjGlb1-2 C79S mutant, shifting to 492 cm^{-1} (Figure 2E), a value indicative of an open heme pocket. Consistently, the amplitude of the geminate recombination becomes $\sim 50\%$, showing that the barrier encountered by the ligand is further decreased. For LjGlb2, the $\nu_{\text{Fe-CO}}$ mode is similar to that observed for CO-ligated mammalian Mbs at neutral pH (Sage et al., 1991). The prevalence of the 5c heme in the deoxy LjGlb2 seems to indicate that, for this globin, the heme-pocket region is indeed more Mb-like.

Plant nsHbs contain Phe(B10) and His(E7) in their distal pockets (Arredondo-Peter et al., 1997; Hoy et al., 2007). These

residues are crucial for protein function because they strongly modulate ligand binding to the heme. The distal His residue in LjGlb1-1 and LjGlb1-2 may also impose a barrier to rebinding and favor ligand escape to the solvent through the protein matrix, thus resulting in small amplitude geminate rebinding. The LjGlb1-1 C8S and C78S mutations do not change substantially the distal pocket interactions of the bound CO, whereas the geminate recombination is changed, suggesting that the distal pocket is somehow perturbed. However, the two mutations appear to lead to opposite effects, as was also the case for the K_H values.

The measurements of NOD activities complemented our spectroscopic studies because NO is a typical ligand of plant and animal Hbs (Brunori et al., 2005; Igamberdiev et al., 2006; Smaghe et al., 2008). Our results demonstrate that both class 1 and class 2 nsHbs are able to scavenge NO at similarly high rates (Figure 6), comparable to those reported for other Hbs using a different assay system (Smaghe et al., 2008). The initial rates of NO consumption measured here are due to genuine NOD activity catalyzed by the hemes and not to Cys nitrosylation because ferric Hbs had no activity and because the wild-type and mutant proteins displayed similar NOD activities. These observations are consistent with the finding that barley Hb1 and its single Cys mutant also showed similar NOD activities (Bykova et al., 2006) and with the proposal that NOD activities are widespread amongst plant and animal Hbs (Smaghe et al., 2008; Gardner, 2012). Interestingly, the NOD activity of LjGlb1-1 would provide the only way, to our knowledge, for this protein to remove bound O₂ because it has an extremely high O₂ affinity ($K_{O_2} = 50$ pM) due to a very slow dissociation rate ($k_{off} = 0.004$ s⁻¹) (Sainz et al., 2013). Because NOD activity yields ferric Hb, enzymatic and/or non-enzymatic systems are required to regenerate oxyferrous Hb and sustain NOD activity *in vivo*. These may include free flavins and flavoproteins (Becana and Klucas, 1992; Igamberdiev et al., 2006; Sanz-Luque et al., 2015). Further studies on the identification of ferric Hb reducing mechanisms shall shed light on this controversial issue (Smaghe et al., 2008).

Our results may be useful for agrobiotechnological applications and for legume researchers in general because they reveal distinct biochemical properties not only between class 1 and class 2 nsHbs, but also between the two members of the same class. The differences in CO binding kinetics found in this work, along with the large variations in O₂ affinities and expression profiles reported earlier (Sainz et al., 2013), strongly suggest that the three proteins perform non-redundant functions. Overexpression of LjGlb1-1 increases the symbiotic performance of *L. japonicus* (Shimoda et al., 2009) and, conversely, the knocked-out line shows alterations in the infection process and produces fewer nodules than the wild-type line (Fukudome et al., 2016). Consequently, transgenic approaches aimed at increasing the content of each of the three nsHbs, or of a combination of them, in the model legume *L. japonicus* (first stage) and in a crop legume with comparable nsHbs, such as soybean or common bean (second stage), are likely to result in outperformance of plants, at least under symbiotic conditions. Likewise, the three genes may be successfully implemented in plant breeding

programs because overexpression of class 1 nsHbs improves survival of hypoxic stress in *A. thaliana* (Hunt et al., 2002) and maize (Mira et al., 2016), and therefore have the potential of conferring tolerance to abiotic stresses.

CONCLUSION

Spectroscopic analyses of LjGlbs reveal major differences between the two phylogenetic classes of nsHbs and also between the two members of the same class, strongly suggesting that the three globins perform non-redundant functions. Specifically, the degree of binding of the distal His(E7) to the heme iron in the deoxyferrous state greatly differs among the LjGlbs studied here. Whereas the equilibrium constant for His binding (K_H) of LjGlb1-1 is in line with those determined for other class 1 nsHbs (Smaghe et al., 2009), LjGlb1-2 behaves more like class 2 nsHbs, showing complete bis-histidyl hexacoordination in the deoxyferrous state. Moreover, LjGlb2 is an atypical class 2 nsHb because it is mostly pentacoordinate in the deoxyferrous form. Upon CO ligation, the bound CO is very strongly stabilized by hydrogen bonding to nearby amino acid residues, probably His(E7), in LjGlb1-1 and its C8S and C78S mutants, but the stabilization is less strong in LjGlb1-2 and LjGlb2. In the latter, the CO is similarly stabilized as in mammalian Mbs and Hbs, which are also pentacoordinate globins. The LjGlb1-1 C8S and C78S mutations caused changes in CO geminate recombination, indicating perturbations of the heme environment. Remarkably, the LjGlb1-2 C79S mutation removes the CO stabilization and gives rise to an open heme pocket. The CO stabilizations of the three globins are consistent with their O₂ affinities (Sainz et al., 2013), following the order LjGlb1-1 > LjGlb1-2 > LjGlb2. Considering that the stronger the stabilization, the higher the affinity, we conclude that the affinities for diatomic ligands are essentially determined by the dissociation rate constants (k_{off}). In contrast to the differences observed for CO binding and geminate recombination, the NOD activities of the three nsHbs were rather similar, which leads us to conclude that the activities are an intrinsic property of the hemes and that the small variations seen in the mutated proteins are due to alterations in the heme environment and not to direct NO scavenging by Cys residues.

AUTHOR CONTRIBUTIONS

All authors performed experiments. SA and SB interpreted results. BC, SVD, SD, CV, and MB interpreted results and wrote parts of the manuscript.

FUNDING

This work was supported by MINECO-Fondo Europeo de Desarrollo Regional (AGL2014-53717-R) and CSIC (Proyecto Intramural Especial 201240E150). SVD, SD, and BC acknowledge the support of the University of Antwerp GOA-BOF funding.

ACKNOWLEDGMENT

LC-B was the recipient of a predoctoral contract (Formación de Personal Investigador) from Ministerio de Economía y Competitividad (MINECO).

REFERENCES

- Abbruzzetti, S., Bruno, S., Faggiano, S., Grandi, E., Mozzarelli, A., and Viappiani, C. (2006). Time-resolved methods in biophysics. 2. Monitoring haem proteins at work with nanosecond laser flash photolysis. *Photochem. Photobiol. Sci.* 5, 1109–1120. doi: 10.1039/b610236k
- Abbruzzetti, S., Grandi, E., Bruno, B., Faggiano, S., Spyraakis, F., Mozzarelli, A., et al. (2007). Ligand migration in nonsymbiotic hemoglobin AHB1 from *Arabidopsis thaliana*. *J. Phys. Chem. B* 111, 12582–12590. doi: 10.1021/jp074954o
- Abbruzzetti, S., Spyraakis, F., Bidon-Chanal, A., Luque, F. J., and Viappiani, C. (2013). Ligand migration through hemeprotein cavities: insights from laser flash photolysis and molecular dynamics simulations. *Phys. Chem. Chem. Phys.* 15, 10686–10701. doi: 10.1039/c3cp51149a
- Andersson, C. R., Ostergaard Jensen, E., Llewellyn, D. J., Dennis, E. S., and Peacock, W. J. (1996). A new hemoglobin gene from soybean: a role for hemoglobin in all plants. *Proc. Natl. Acad. Sci. U.S.A.* 93, 5682–5687. doi: 10.1073/pnas.93.12.5682
- Appleby, C. A. (1984). Leghemoglobin and rhizobium respiration. *Annu. Rev. Plant Physiol.* 35, 443–478. doi: 10.1146/annurev.pp.35.060184.002303
- Appleby, C. A., Blumberg, W. E., Peisach, J., Wittenberg, B. A., and Wittenberg, J. B. (1976). Leghemoglobin. An electron paramagnetic resonance and optical spectral study of the free protein and its complexes with nicotinate and acetate. *J. Biol. Chem.* 251, 6090–6096.
- Arredondo-Peter, R., Hargrove, M. S., Sarath, G., Moran, J. F., Lohrman, J., Olson, J. S., et al. (1997). Rice hemoglobins. Gene cloning, analysis, and O₂-binding kinetics of a recombinant protein synthesized in *Escherichia coli*. *Plant Physiol.* 115, 1259–1266. doi: 10.1104/pp.115.3.1259
- Becana, M., and Klucas, R. V. (1992). Oxidation and reduction of leghemoglobin in root nodules of leguminous plants. *Plant Physiol.* 98, 1217–1221. doi: 10.1104/pp.98.4.1217
- Bisht, N. K., Abbruzzetti, S., Uppal, S., Bruno, S., Spyraakis, F., Mozzarelli, A., et al. (2011). Ligand migration and hexacoordination in type 1 non-symbiotic rice hemoglobin. *Biochim. Biophys. Acta* 1814, 1042–1053. doi: 10.1016/j.bbapap.2010.09.016
- Bogusz, D., Appleby, C. A., Landsmann, J., Dennis, E. S., Trinick, M. J., and Peacock, W. J. (1988). Functioning haemoglobin genes in non-nodulating plants. *Nature* 331, 178–180. doi: 10.1038/331178a0
- Bruno, S., Faggiano, S., Spyraakis, F., Mozzarelli, A., Abbruzzetti, S., Grandi, E., et al. (2007a). The reactivity with CO of AHB1 and AHB2 from *Arabidopsis thaliana* is controlled by the distal HisE7 and internal hydrophobic cavities. *J. Am. Chem. Soc.* 129, 2880–2889.
- Bruno, S., Faggiano, S., Spyraakis, F., Mozzarelli, A., Cacciatori, E., Dominici, P., et al. (2007b). Different roles of protein dynamics and ligand migration in non-symbiotic hemoglobins AHB1 and AHB2 from *Arabidopsis thaliana*. *Gene* 398, 224–233.
- Brunori, M. (2001). Nitric oxide moves myoglobin centre stage. *Trends Biochem. Sci.* 26, 209–210. doi: 10.1016/S0968-0004(01)01824-2
- Brunori, M., Giuffrè, A., Nienhaus, K., Nienhaus, G. U., Scandurra, F. M., and Vallone, B. (2005). Neuroglobin, nitric oxide, and oxygen: functional pathways and conformational changes. *Proc. Natl. Acad. Sci. U.S.A.* 102, 8483–8488. doi: 10.1073/pnas.0408766102
- Bustos-Sanmamed, P., Tovar-Méndez, A., Crespi, M., Sato, S., Tabata, S., and Becana, M. (2011). Regulation of nonsymbiotic and truncated hemoglobin genes of *Lotus japonicus* in plant organs and in response to nitric oxide and hormones. *New Phytol.* 189, 765–776. doi: 10.1111/j.1469-8137.2010.03527.x
- Bykova, N. V., Igamberdiev, A. U., Ens, W., and Hill, R. D. (2006). Identification of an intermolecular disulfide bond in barley hemoglobin. *Biochem. Biophys. Res. Commun.* 347, 301–309. doi: 10.1016/j.bbrc.2006.06.091
- Cerda-Colon, J. F., Silfa, E., and Lopez-Garriga, J. (1998). Unusual rocking freedom of the heme in the hydrogen sulfide-binding hemoglobin from *Lucina pectinata*. *J. Am. Chem. Soc.* 120, 9312–9317. doi: 10.1021/ja972654m
- Das, T. K., Lee, H. C., Duff, S. M. G., Hill, R. D., Peisach, J., Rousseau, D. L., et al. (1999). The heme environment in barley hemoglobin. *J. Biol. Chem.* 274, 4207–4212. doi: 10.1074/jbc.274.7.4207
- Duff, S. M. G., Wittenberg, J. B., and Hill, R. D. (1997). Expression, purification, and properties of recombinant barley (*Hordeum* sp.) hemoglobin. *J. Biol. Chem.* 272, 16746–16752. doi: 10.1074/jbc.272.27.16746
- Fukudome, M., Calvo-Begueria, L., Kado, T., Osuki, K., Rubio, M. C., Murakami, E., et al. (2016). Hemoglobin LjGlb1-1 is involved in nodulation and regulates the level of nitric oxide in the *Lotus japonicus*-*Mesorhizobium loti* symbiosis. *J. Exp. Bot.* 67, 5275–5283. doi: 10.1093/jxb/erw290
- Gardner, P. R. (2012). Hemoglobin: a nitric-oxide dioxygenase. *Scientifica* 2012, 1–34. doi: 10.6064/2012/683729
- Goodman, M. D., and Hargrove, M. S. (2001). Quaternary structure of rice nonsymbiotic hemoglobin. *J. Biol. Chem.* 276, 6834–6839. doi: 10.1074/jbc.M009254200
- Handberg, K., and Stougaard, J. (1992). *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant J.* 2, 487–496. doi: 10.1111/j.1365-313X.1992.00487.x
- Hebelstrup, K. H., and Jensen, E. Ø. (2008). Expression of NO scavenging hemoglobin is involved in the timing of bolting in *Arabidopsis thaliana*. *Planta* 227, 917–927. doi: 10.1007/s00425-007-0667-z
- Hoy, J. A., Robinson, H., Trent, I. I. J. T., Kakar, S., Smaghe, B. J., and Hargrove, M. S. (2007). Plant hemoglobins: a molecular fossil record for the evolution of oxygen transport. *J. Mol. Biol.* 371, 168–179. doi: 10.1016/j.jmb.2007.05.029
- Hu, S., Smith, K. M., and Spiro, T. G. (1996). Assignment of protoheme Resonance Raman spectrum by heme labeling in myoglobin. *J. Am. Chem. Soc.* 118, 12638–12646. doi: 10.1021/ja962239e
- Hunt, P. W., Klok, E. J., Trevaskis, B., Watts, R. A., Ellis, M. H., Peacock, W. J., et al. (2002). Increased level of hemoglobin 1 enhances survival of hypoxic stress and promotes early growth in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17197–17202. doi: 10.1073/pnas.212648799
- Hunt, P. W., Watts, R. A., Trevaskis, B., Llewellyn, D. J., Burnell, J., Dennis, E. S., et al. (2001). Expression and evolution of functionally distinct haemoglobin genes in plants. *Plant Mol. Biol.* 47, 677–692. doi: 10.1023/A:1012440926982
- Igamberdiev, A. U., Bykova, N. V., and Hill, R. D. (2006). Nitric oxide scavenging by barley hemoglobin is facilitated by a monodehydroascorbate reductase-mediated ascorbate reduction of methemoglobin. *Planta* 223, 1033–1040. doi: 10.1007/s00425-005-0146-3
- Igamberdiev, A. U., and Hill, R. D. (2004). Nitrate, NO and haemoglobin in plant adaptation to hypoxia: an alternative to classic fermentation pathways. *J. Exp. Bot.* 55, 2473–2482. doi: 10.1093/jxb/erh272
- Ioanitescu, A. I., Dewilde, S., Kiger, L., Marden, M. C., Moens, L., and Van Doorslaer, S. (2005). Characterization of nonsymbiotic tomato hemoglobin. *Biophys. J.* 89, 2628–2639. doi: 10.1529/biophysj.105.060582
- Kakar, S., Hoffman, F. G., Storz, J. F., Fabian, M., and Hargrove, M. S. (2010). Structure and reactivity of hexacoordinate hemoglobins. *Biophys. Chem.* 152, 1–14. doi: 10.1016/j.bpc.2010.08.008
- Mira, M. M., Hill, R. D., and Stasolla, C. (2016). Phytoglobins improve hypoxic root growth by alleviating apical meristem cell death. *Plant Physiol.* 172, 2044–2056. doi: 10.1104/pp.16.01150
- Perazzolli, M., Dominici, P., Romero-Puertas, M. C., Zago, E., Zeier, J., Sonoda, M., et al. (2004). Nonsymbiotic hemoglobin AHB1 modulates nitric oxide bioactivity in *Arabidopsis thaliana*. *Plant Cell* 16, 2785–2794. doi: 10.1105/tpc.104.025379
- Rousseau, D. L., Ondrias, M. R., LaMar, G. N., Kong, S. B., and Smith, K. M. (1983). Resonance Raman spectra of the heme in leghemoglobin. Evidence for

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00407/full#supplementary-material>

- the absence of ruffling and the influence of the vinyl groups. *J. Biol. Chem.* 258, 1740–1746.
- Sage, J. T., Morikis, D., and Champion, P. M. (1991). Spectroscopic studies of myoglobin at low pH: heme structure and ligation. *Biochemistry* 30, 1227–1237. doi: 10.1021/bi00219a010
- Sainz, M., Pérez-Rontomé, C., Ramos, J., Mulet, J. M., James, E. K., Bhattacharjee, U., et al. (2013). Plant hemoglobins may be maintained in functional form by reduced flavins in the nuclei, and confer differential tolerance to nitro-oxidative stress. *Plant J.* 76, 875–887. doi: 10.1111/tpj.12340
- Sanz-Luque, E., Ocaña-Calahorra, F., de Montaigne, A., Chamizo-Ampudia, A., Llamas, A., Galván, A., et al. (2015). THB1, a truncated hemoglobin, modulates nitric oxide levels and nitrate reductase activity. *Plant J.* 81, 467–479. doi: 10.1111/tpj.12744
- Sasakura, F., Uchiyama, T., Shimoda, Y., Suzuki, A., Takenouchi, K., Higashi, S., et al. (2006). A class 1 hemoglobin gene from *Alnus firma* functions in symbiotic and nonsymbiotic tissues to detoxify nitric oxide. *Mol. Plant Microbe Interact.* 19, 441–450. doi: 10.1094/MPMI-19-0441
- Shimoda, Y., Shimoda-Sasakura, F., Kucho, K., Kanamori, N., Nagata, M., Suzuki, A., et al. (2009). Overexpression of class 1 plant hemoglobin genes enhances symbiotic nitrogen fixation activity between *Mesorhizobium loti* and *Lotus japonicus*. *Plant J.* 57, 254–263. doi: 10.1111/j.1365-313X.2008.03689.x
- Smaghe, B. J., Hoy, J. A., Percifield, R., Kundu, S., Hargrove, M. S., Sarath, G., et al. (2009). Correlations between oxygen affinity and sequence classifications of plant hemoglobins. *Biopolymers* 91, 1083–1096. doi: 10.1002/bip.21256
- Smaghe, B. J., Sarath, G., Ross, E., Hilbert, J., and Hargrove, M. S. (2006). Slow ligand binding kinetics dominate ferrous hexacoordinate hemoglobin reactivities and reveal differences between plants and other species. *Biochemistry* 45, 561–570. doi: 10.1021/bi051902l
- Smaghe, B. J., Trent, I. I. J. T., and Hargrove, M. S. (2008). NO dioxygenase activity in hemoglobins is ubiquitous in vitro, but limited by reduction in vivo. *PLoS ONE* 3:e2039. doi: 10.1371/journal.pone.0002039
- Spiro, T. G., and Wasbotten, I. H. (2005). CO as a vibrational probe of heme protein active sites. *J. Inorg. Biochem.* 99, 34–44. doi: 10.1016/j.jinorgbio.2004.09.026
- Spyrakakis, F., Luque, C. F. J., and Viappiani, C. (2011). Structural analysis in nonsymbiotic hemoglobins: what can we learn from inner cavities? *Plant Sci.* 181, 8–13. doi: 10.1016/j.plantsci.2011.03.021
- Taylor, E. R., Nie, X. Z., MacGregor, A. W., and Hill, R. D. (1994). A cereal haemoglobin gene is expressed in seed and root tissues under anaerobic conditions. *Plant Mol. Biol.* 24, 853–862. doi: 10.1007/BF00014440
- Tjepkema, J. D., Schwintzer, C. R., and Benson, D. R. (1986). Physiology of actinorhizal nodules. *Annu. Rev. Plant Physiol.* 37, 209–232. doi: 10.1146/annurev.pp.37.060186.001233
- Trent, I. I. J. T., Watts, R. A., and Hargrove, M. S. (2001). Human neuroglobin, a hexacoordinate hemoglobin that reversibly binds oxygen. *J. Biol. Chem.* 276, 30106–30110. doi: 10.1074/jbc.C100300200
- Trevaskis, B., Watts, R. A., Andersson, C. R., Llewellyn, D. J., Hargrove, M. S., Olson, J. S., et al. (1997). Two hemoglobin genes in *Arabidopsis thaliana*: the evolutionary origins of leghemoglobins. *Proc. Natl. Acad. Sci. U.S.A.* 94, 12230–12234. doi: 10.1073/pnas.94.22.12230
- Uzan, J., Dewilde, S., Burmester, T., Hankeln, T., Moens, L., Hamdane, D., et al. (2004). Neuroglobin and other hexacoordinated hemoglobins show a weak temperature dependence of oxygen binding. *Biophys. J.* 87, 1196–1204.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Calvo-Begueria, Cuypers, Van Doorslaer, Abbruzzetti, Bruno, Berghmans, Dewilde, Ramos, Viappiani and Becana. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Determination of Photoperiod-Sensitive Phase in Chickpea (*Cicer arietinum* L.)

Ketema Daba¹, Thomas D. Warkentin¹, Rosalind Bueckert¹, Christopher D. Todd² and Bunyamin Tar'an^{1*}

¹ Crop Development Centre/Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada,

² Department of Biology, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:

Maria Carlota Vaz Patto,
Universidade Nova de Lisboa,
Portugal

Reviewed by:

Roberta Paradiso,
University of Naples Federico II, Italy
Ivan A. Matus,
Instituto de Investigaciones
Agropecuarias, Chile

*Correspondence:

Bunyamin Tar'an
bunyamin.taran@usask.ca

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 03 December 2015

Accepted: 24 March 2016

Published: 11 April 2016

Citation:

Daba K, Warkentin TD, Bueckert R,
Todd CD and Tar'an B (2016)
Determination
of Photoperiod-Sensitive Phase
in Chickpea (*Cicer arietinum* L.).
Front. Plant Sci. 7:478.
doi: 10.3389/fpls.2016.00478

Photoperiod is one of the major environmental factors determining time to flower initiation and first flower appearance in plants. In chickpea, photoperiod sensitivity, expressed as delayed to flower under short days (SD) as compared to long days (LD), may change with the growth stage of the crop. Photoperiod-sensitive and -insensitive phases were identified by experiments in which individual plants were reciprocally transferred in a time series from LD to SD and vice versa in growth chambers. Eight chickpea accessions with differing degrees of photoperiod sensitivity were grown in two separate chambers, one of which was adjusted to LD (16 h light/8 h dark) and the other adjusted to SD (10 h light/14 h dark), with temperatures of 22/16°C (12 h light/12 h dark) in both chambers. The accessions included day-neutral (ICCV 96029 and FLIP 98-142C), intermediate (ICC 15294, ICC 8621, ILC 1687, and ICC 8855), and photoperiod-sensitive (CDC Corinne and CDC Frontier) responses. Control plants were grown continuously under the respective photoperiods. Reciprocal transfers of plants between the SD and LD photoperiod treatments were made at seven time points after sowing, customized for each accession based on previous data. Photoperiod sensitivity was detected in intermediate and photoperiod-sensitive accessions. For the day-neutral accession, ICCV 96029, there was no significant difference in the number of days to flowering of the plants grown under SD and LD as well as subsequent transfers. In photoperiod-sensitive accessions, three different phenological phases were identified: a photoperiod-insensitive pre-inductive phase, a photoperiod-sensitive inductive phase, and a photoperiod-insensitive post-inductive phase. The photoperiod-sensitive phase extends after flower initiation to full flower development. Results from this research will help to develop cultivars with shorter pre-inductive photoperiod-insensitive and photoperiod-sensitive phases to fit to regions with short growing seasons.

Keywords: adaptation, flowering, photoperiod-insensitive phase, long days, short days

INTRODUCTION

In western Canada, the short crop growing season available for chickpea (110–120 days) often coincides with end-of-season frost resulting in severe losses in grain yield and quality (Warkentin et al., 2003). In order to maximize crop yield through agronomic management or plant breeding, the phenology of the crop must be well matched to the resources and constraints of the production

Abbreviations: LD, long day; PS, photoperiod sensitivity; SD, short day.

environment (Summerfield et al., 1996; Levy and Dean, 1998; Warkentin et al., 2003). All flowering plants undergo several developmental transitions during their life cycle which can be divided into three major physiological developmental phases: a vegetative phase, from emergence to flower initiation, the reproductive phase, from floral initiation to anthesis, and physiological maturity, from anthesis to seed filling (Ritchie, 1991; Ritchie et al., 1998). The vegetative growth phase is comprised of the basic vegetative phase and the photoperiod-sensitive phase (Vergara and Chang, 1985).

The transition from the vegetative to reproductive phase is a major developmental switch in the plant's life cycle (Levy and Dean, 1998). This transition is crucial for survival because plants normally time the onset of flowering to suitable environmental conditions. Many plant species have evolved the ability to initiate flowering in response to environmental factors such as changes in photoperiod and temperature. The beginning stage of flowering commands the start of the seed-set period, and, thus, is a key stage in yield formation (Lejeune-Hénaut et al., 1999; Putterill et al., 2004). Flower development and the seed-set stages are greatly impeded by stress such as drought and frost, so flowering and seed development must be completed during favorable growing conditions. Timely flowering and maturity in relation to the available growing season in a particular location are essential for high yield potential from annual crops (Bunting, 1975). Understanding the photoperiod-sensitive phase of a photoperiodic plant would allow better crop management strategy to either promote early flowering to reduce crop duration time, or to intentionally delay flowering (Warner, 2009).

In experiments on rice (*Oryza sativa* L.), where transfers were made between long days (LD) and short days (SD) and vice versa, the photoperiod-sensitive phase was flanked by two photoperiod-insensitive phases (Yin, 2008). In wheat (*Triticum aestivum* L.), a LD plant, the full pre-anthesis period was found to be divided into three sub-phases: from sowing to the terminal spikelet, from terminal spikelet initiation to heading, and from heading to anthesis indicating stage-dependence of plant responsiveness to temperature (Slafer and Rawson, 1995). In this crop, the flowering response was affected by temperature throughout their life cycles (Slafer and Rawson, 1994). In SD crop species such as cowpea (*Vigna unguiculata* L. Walp.) and soybean (*Glycine max* L. Merr.), there was a temperature-dependent critical photoperiod phase. Beyond the critical point, time to flowering was solely a function of mean temperature (Hadley et al., 1984). Maize (*Zea mays* L.) is also sensitive to photoperiod during tassel initiation (Kiniry et al., 1983). In rice and soybean, the photoperiod influence even extends for some time beyond the phase of floral initiation (Collinson et al., 1992; Ellis et al., 1992).

Earlier studies on wheat reported that exposure to long photoperiod significantly reduced the time to heading (Slafer and Rawson, 1994, 1997). Estimation of phasic development is crucial for accurate modeling of plant development and yield components, as well as for evaluating cultivar adaptation and scheduling cultural practices (Shaykewich, 1995). Quantitative models to determine the developmental phases in different

plants were developed using different parameters and plant materials. Flower development phases were quantified using four parameters; a_1 (the photoperiod-insensitive pre-inductive phase), I_s (the photoperiod-sensitive inductive phase in LD and SD), and a_3 the photoperiod-insensitive post-inductive phases in LD and SD (Ellis et al., 1992). Similarly, photoperiod-sensitive inductive phases in LD and SD were denoted as I_{2L} and I_{2S} , respectively, by following the procedure developed by Yin (2008).

Short-day does not delay flowering in a LD plant if exposure is restricted to the photoperiod-insensitive pre-inductive phase or the photoperiod-insensitive phase of flower development. However, time to flower is delayed if the plant is exposed to SD during the photoperiod-sensitive phase. Similarly, LD will only hasten flowering in LD plants if the plants are exposed to the photoperiod when they are at the photoperiod-sensitive stage as summarized by Adams et al. (2001, 2003). The duration of the photoperiod-sensitive phases can be determined by examining data on the time to first flower opening of plants transferred between SD and LD at different times (Wang et al., 1997; Yin et al., 1997; Adams et al., 2001). Chickpea is inherently considered as a LD plant (Soltani et al., 2004). Chickpea accessions with day neutral, intermediate, and highly sensitive response to photoperiod were recently reported by Daba et al. (2015). Little is known about the duration of the photoperiod-sensitive and -insensitive phases in chickpea. Therefore, the reciprocal transfer technique was used to quantify and identify the timing and duration of the photoperiod-sensitive phase and the time of floral initiation in chickpea. The objectives of this research were to determine the timing and duration of the photoperiod-sensitive and photoperiod-insensitive phase in selected chickpea accessions representative of different maturity classes, and to establish whether photoperiod sensitivity ends at floral initiation or extends into the phase of flower development.

MATERIALS AND METHODS

Accessions Evaluated

Eight diverse chickpea accessions namely: ICCV 96029 (S1), FLIP 98-142C (S2), ICC 15294 (I1), ICC 8621 (I2), ILC 1687 (I3), and ICC 8855 (I4), CDC Corinne (S1), and CDC Frontier (S2) collected from the gene banks of the International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), India and the International Center for Agricultural Research in the Dry Areas (ICARDA), together with cultivars developed at the Crop Development Centre, University of Saskatchewan were used in this research (Table 1). These eight genotypes are referred to as 'accessions' throughout this paper.

The accessions were grown in two separate growth chambers: one of the chambers was adjusted to SD of 10 h light period (SD) and the other chamber was adjusted to LD of 16 h light period (LD). The chambers were maintained at day/night temperatures of 22/16°C (12/12 h). The 12/12 h cycle was used to avoid confounding effects of asynchrony between thermal and photoperiod factors (Roberts et al., 1985; Yin, 2008). Both growth

TABLE 1 | Market class, origin, and potential photoperiod sensitivity group of chickpea accessions used in the determination of photoperiod sensitivity phase by reciprocally transferring plants from LD to SD.

Accessions	Market class	Origin	Difference in number of days to flowering (SD–LD)	Photoperiod sensitivity group
CDC Corinne	Desi	CDC, University of Saskatchewan	62	S1
CDC Frontier	Kabuli	CDC, University of Saskatchewan	60	S2
ICC 15294	Desi	Iran	42	I1
ICC 8621	Desi	Ethiopia	20	I2
ILC 1687	Kabuli	Afghanistan	27	I3
ICC 8855	Kabuli	Afghanistan	40	I4
ICCV 96029	Desi	ICRISAT	2	N1
FLIP 98-142C	Kabuli	ICARDA	10	N2

chambers were equipped with inflorescent light bulbs with total light irradiance of $370 \mu\text{mol m}^{-2}\text{s}^{-1}$ at just above the plant canopy.

Three seeds of each accession were planted in 3.8 L pots containing Sunshine mix #4 (Sun Gro, Seba Beach, AB, Canada). Seedlings were thinned to two plants per pot 2 weeks after sowing or after full emergence of the seedlings. Starting from 1 week after crop emergence, the plants were watered every 2–3 days based on the growth stage and water use of each accession. Once a week a quick release fertilizer (20 N:20 P_2O_5 :20 K_2O) prepared at a concentration of 3 g L^{-1} was applied at a rate of 100 ml per pot starting 1 week after emergence.

A total of three pots with two plants per pot were assigned for each treatment. Each transfer treatment, therefore, contained a total of six plants. A total of six pots were also assigned as a control for each accession. Within the growth chambers, a total of 27 pots, for each of the seven transfers and the control pots for each accession were completely randomized. Control plants were continuously grown at LD and SD (Table 2 and Figure 1). Transfer times for these accessions were customized based on their differences in the number of days to flowering under short compared to LD (Daba et al., 2015). Once plants had been transferred, they were continuously maintained in the new chamber under either LD or SD. The entire experiment was carried out in two runs (two time replicates).

TABLE 2 | Chickpea accessions used in the determination of photoperiod-sensitivity phase and days from sowing to transfer (tc) for plants moved from LD to SD and vice versa.

Accessions	Transfer times (days after sowing)							
CDC Corinne (S1)	0	11	22	33	44	55	66	77
CDC Frontier (S2)	0	10	20	30	40	50	60	70
ICC 15294 (I1)	0	9	18	27	36	45	54	63
ICC 8621 (I2)	0	8	16	24	32	40	48	56
ILC 1687 (I3)	0	8	16	24	32	40	48	56
ICC 8855 (I4)	0	6	12	18	24	30	36	42
ICCV 96029 (N1)	0	5	10	15	20	25	30	35
FLIP 98-142C (N2)	0	5	10	15	20	25	30	35

0 day to transfer refers to control plants which were grown under SD or LD throughout.

Data Collection and Analysis

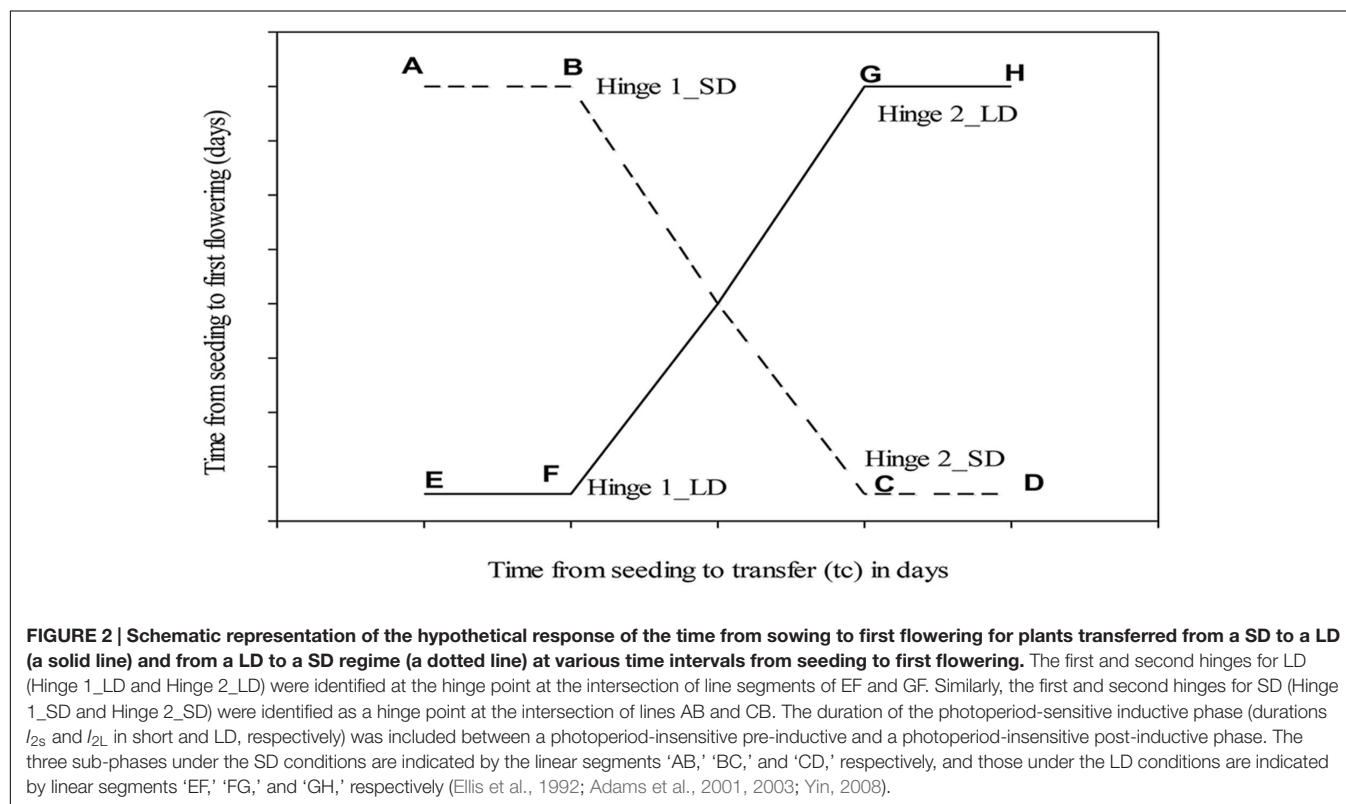
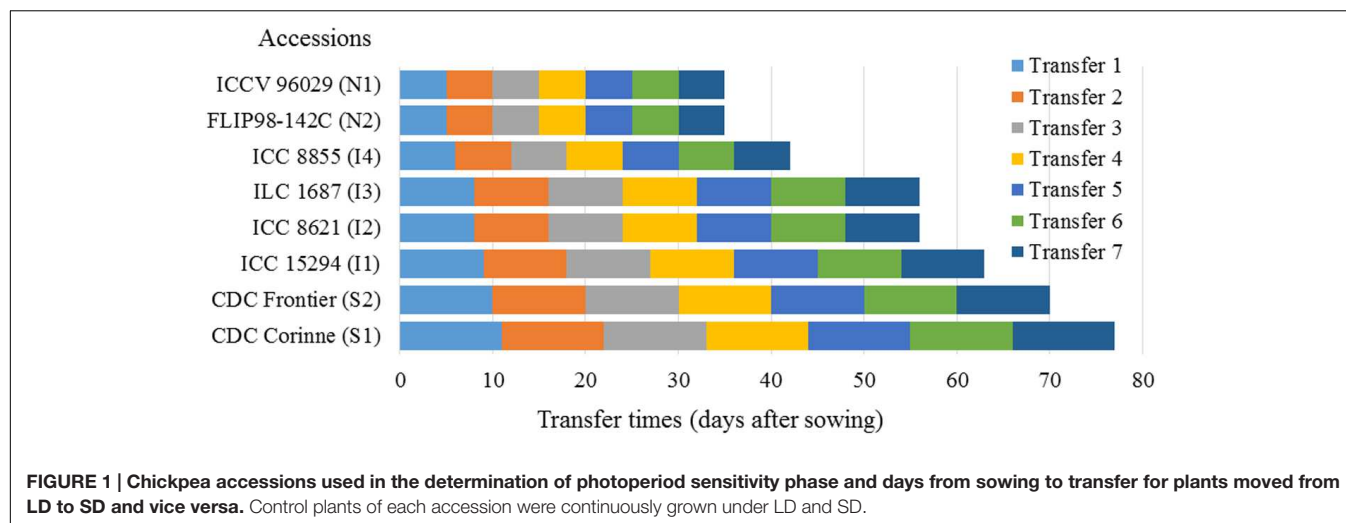
Flower Bud Initiation and Full Flower Opening

First flower bud initiation stage and full open flower appearance (corolla visible) were recorded for each accession. Samples of flower buds were carefully collected from each control and transferred plants in both photoperiod treatments. The stipules were dissected using blades to expose the shoot apex and newly initiated phytomers and a node subtending a leaf primordium, and an auxiliary vegetative or reproductive bud and were evaluated under the microscope. In cases of the death of the first initiated buds, the subsequently formed buds were dissected. Upon seeing fully developed anthers bound by a fully developed calyx, the flower bud initiation stages (as the number of days to flower bud initiation from days of seeding) were declared. Days to first flowering for the transfer and control plants in LD and SD were recorded when a fully opened flower appeared on each plant. Days to flower bud initiation and first flowering of the control plants in short and LD photoperiods were used for comparison.

Identification of Slope Coefficients, y-Intercepts and Hinges

Segmented linear regression analysis was conducted for each accession in order to determine the differences in the photoperiod-sensitive and photoperiod-insensitive phases in the chickpea accessions reciprocally transferred from LD to SD. The hypothetical response of the time from sowing to first flowering for plants transferred from a SD to a LD and from a LD to a SD regime at various time intervals from seeding to first flowering were illustrated in Figure 2. Control plants continuously grown under SD are indicated by point A, and those grown under LD are indicated by point E. The intersection point of linear segments AB and CB representing the first hinge point for transfer from LD to SD, whereas the intersection point between linear segments of EF and FG represented the first hinge for transfers from SD to LD. Accordingly, the first hinge was calculated as a function of days of transfer from seeding to days to flowering from seeding where the increase or decrease in the slope between the linear segments occurred.

The two-phase or piecewise regression (hinged regression) has been described by Breiman (1993) and used by Dunnigan

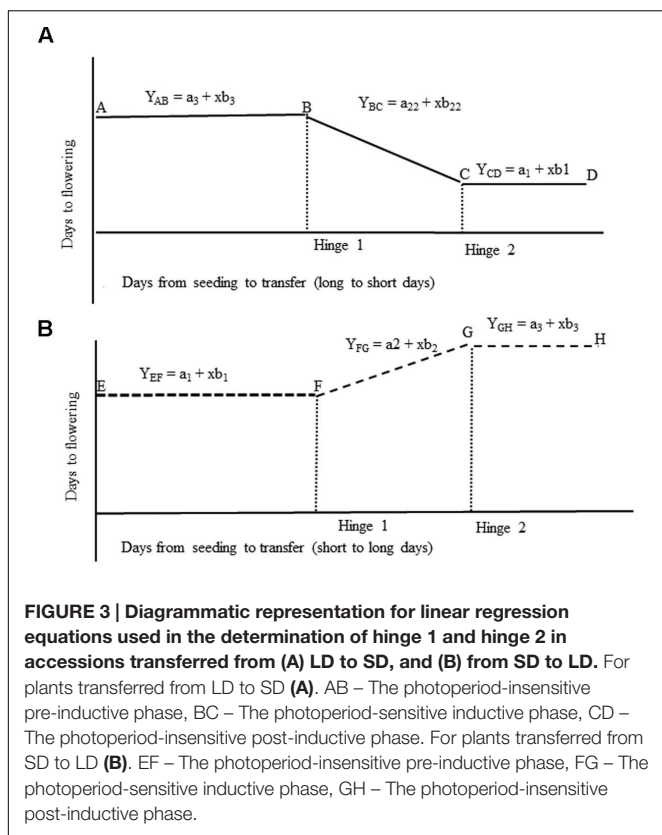


et al. (1997), and Kenesei and Abonyi (2013). The individual linear segments were then used to determine the photoperiod-sensitive and -insensitive phases in chickpea accessions following a procedure described by Wang et al. (1997). The intersection of the two linear equations, the hinge, was positioned to identify the changes in slope coefficients and y-intercepts using simultaneous equations for each part of the regression models (Figure 3). The models were interactively congregated in PROC MODEL of SAS version 9.3 (SAS Institute Inc., 2009; SAS Institute Inc., Cary, NC, USA). Hinged regression analyses were conducted for each accession to determine the parameters (a_1 , b_1 , a_2 , b_2 , hinge 1,

a_{22} , b_{22} , a_3 , b_3 , and hinge 2) where a and b represent intercept and slope coefficients of the respective segments of the modeled line. To test if regression intercept and slope coefficient varied among accessions, analyses of variance were conducted using PROC GLM of SAS version 9.3 (SAS Institute Inc., 2009; SAS Institute Inc., Cary, NC, USA).

Identification of Photoperiod-Sensitive and Photoperiod-Insensitive Phases

Data on days to flowering for the transfer and control plants (Figure 4) were analyzed using PROC NON-LINEAR of SAS



version 9.3 (SAS Institute Inc., 2009; SAS Institute Inc., Cary, NC, USA). Initially, separate data analyses were conducted for each time replicate. There was no significant difference between the results of the two time replicates. Homogeneity of variance for each time replicate was validated using Levene's Test. Thus, a combined data analysis was conducted using the average data of the replications in both time replicates for each accession transferred from LD to SD and vice versa, and the control plants.

The time to flower in the eight chickpea accessions in the successive photoperiod-transfer and control treatments were modeled against the time to transfer after seeding following the procedure of Yin (2008) as illustrated in Figure 2, which combines data from transfers of LD to SD and short to LD in a single curve fitting procedure.

In the analysis, f_L was assigned as the duration from sowing to flowering for the LD and can be written as: $f_L = I_{1L} + I_{2L} + I_{3L}$; where: I_{1L} is the first sub-phase, a photoperiod-insensitive pre-inductive phase; I_{2L} is the second sub-phase, a photoperiod-sensitive inductive phase; I_{3L} is the third sub-phase, a photoperiod-insensitive post-inductive phase under LD conditions (Yin, 2008). Similarly f_S was assigned as the duration from sowing to flowering for the SD and the expression can be written as: $f_S = I_{1S} + I_{2S} + I_{3S}$; where: I_{1S} is the first sub-phase a photoperiod-insensitive pre-inductive phase; I_{2S} is the second sub-phase a photoperiod-sensitive phase; I_{3S} is the third sub-phase, a photoperiod-insensitive post-inductive phase under SD conditions.

RESULTS

Flower Bud Initiation and Full Flower Opening

The average number of days to flowering was lower for control plants grown continuously under LD compared to those under SD (Table 3). In the photoperiod-sensitive accessions, [CDC Corinne (S1) and CDC Frontier (S2)], flowering time was delayed by 45 and 38 days, respectively, under SD compared to LD. Delay in flowering of the four accessions with intermediate response to photoperiod [ICC 15294 (I1), ICC 8621 (I2), ILC1867 (I3), and ICC 8855 (I4)] ranged from 17 to 42 days under SD compared to LD. Flowering in the day-neutral accessions, [ICCV 96029 (N1) and FLIP 98-142C (N2)], was delayed by 1 and 10 days, respectively, under SD as compared to LD.

Slope Coefficients, y-Intercepts and Hinges

The slope coefficients, y-intercepts and the first and second hinges as determined by the simultaneous linear equations are listed in (Table 4). There were significant differences among the chickpea accessions for the second hinge, intercepts (a_1 , a_{22} , and a_3) and slope coefficients (b_1 , b_2 , and b_{22} ; $P \leq 0.0001$). However, the difference among the accessions for the first hinge, the initial intercept (a_2), and the slope coefficients (b_3) of the simultaneous equations were not significant (Table 5).

In our analysis, the first hinge corresponds to a beginning of change in time from seeding to first flowering against time from seeding to transfer. The photoperiod-sensitive accessions had the highest values of both the first and second hinge values. The intermediate accessions had intermediate values of both first and second hinges. The values of the first and second hinges for the day-neutral accession, ICCV 96029 (N1) were identified to be 0. The identified hinges facilitated determination of photoperiod-sensitive phase in chickpea accessions. The difference between hinge 1 and hinge 2 was considered as the photoperiod-sensitive phase. Accordingly, in CDC Frontier (S2), a photoperiod-sensitive accession, the first hinge and second hinge were 16 and 47 days, respectively. Based on the difference between the first and the second hinges, 31 days was considered as the length of the photoperiod-sensitive phase of this accession. Similarly, CDC Corinne (S1) and ICC 15294 (I1) each had the first hinge value of 20 days. The values of the second hinge for these two accessions were 42 and 45 days, respectively. The duration of photoperiod sensitivity of these accessions based on the difference between the second and first hinge were 22 and 25 days, respectively. For other intermediate accessions, the values of the first hinge were 9 to 10 days. The second hinge for these accessions was 19 days. Thus the duration of the photoperiod-sensitive phase ranged from 15 to 26 days.

Linear Regression

The slope coefficient values of the accessions were negative for transfers from long to SD ranging from -0.40 to -1.00 (Table 6). On the other hand, the slope values of the accessions transferred from short to LD were positive ranging from

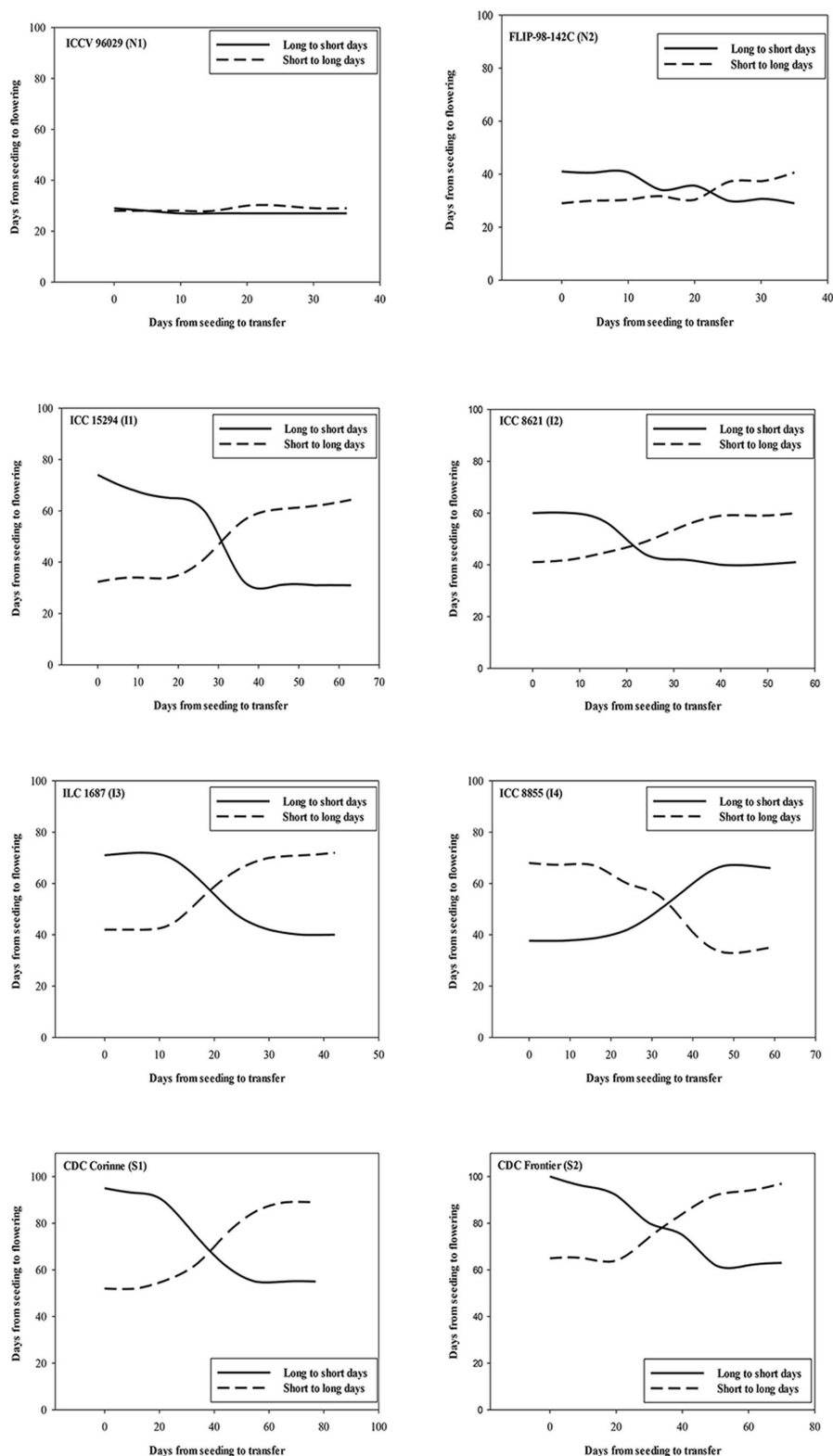


FIGURE 4 | The effect of transferring the plants at varying intervals from LD to SD (solid line) and SD to LD (dashed line) on the number of days to first flower opening for each of day neutral (ICCV 96029, FLIP 98-142C), intermediate (ICC 15294, ICC 8621, ILC 1687, and ICC 8855) and photoperiod-sensitive (CDC Corinne and CDC Frontier) accessions.

TABLE 3 | Average number of days from seeding to first flower bud initiation and the number of days to flower under long and SD photoperiod conditions over two time replicates of the experiment.

Accessions	Days from seeding to flower bud initiation		Days from seeding to first flowering	
	LD \pm s.d.	SD \pm s.d.	LD \pm s.d.	SD \pm s.d.
CDC Corinne (S1)	38 \pm 1.0	39 \pm 1.0	63 \pm 2.5	101 \pm 1.6
CDC Frontier (S2)	30 \pm 1.0	38 \pm 0.0	60 \pm 3.3	105 \pm 0.9
ICC 15294 (I1)	28 \pm 1.0	31 \pm 1.0	32 \pm 0.5	74 \pm 2.9
ICC 8621 (I2)	30 \pm 0.0	30 \pm 1.2	38 \pm 1.3	55 \pm 4.1
ILC 1687 (I3)	28 \pm 1.0	31 \pm 0.0	41 \pm 1.3	66 \pm 0.0
ICC 8855 (I4)	29 \pm 0.0	30 \pm 1.0	38 \pm 1.3	68 \pm 1.6
ICCV 96029 (N1)	19 \pm 0.0	19 \pm 0.0	28 \pm 0.6	29 \pm 0.0
FLIP 98-142C (N2)	22 \pm 0.0	24 \pm 0.0	30 \pm 0.0	40 \pm 0.8

0.05 to 0.95. The slopes for ICCV 96029 (N1) transferred from LD to SD and from SD to LD were 0 and 0.05, respectively.

Photoperiod-Sensitive and Photoperiod-Insensitive Phases in Chickpea Accessions

The reciprocal transfer model fitted the data with R^2 -values among the accessions ranging from 0.74 to 0.99 (Table 7). Three developmental phases were identified in all the accessions except ICCV 96029 (N1).

Photoperiod-Insensitive Pre-inductive Phase

In the photoperiod-sensitive accessions, a photoperiod-insensitive pre-inductive phase of 15–19 days was observed under LD, and 17–20 days under SD. These values ranged from 9 to 13 days in the intermediate accessions under LD and from 13 to 18 days under SD. In ICCV 96029 (N1), the values of the photoperiod-insensitive pre-inductive phase were 22 and 23 days under long and SD, respectively. This value ranged between 18 and 20 days under long and SD, respectively, in FLIP 98-142C (N2).

TABLE 5 | Analysis of variance for a1, b1, a2, b2, hinge 1, a22, b22, a3, b3, and hinge 2 of the eight chickpea accessions used in the photoperiod-sensitive and -insensitive phase determination using reciprocal transfers between SD (10 h light) and LD (16 h light) photoperiod conditions over two time replicates.

Characters	Source of variation			
	df	Sum of squares	Mean square	F-value
a1	7	10995	1571	278***
b1	7	22342	3192	13***
a2	7	0.2	0	1.3 ^{ns}
b2	7	14	2	6.6**
Hinge 1	7	805	115	2.9 ^{ns}
a22	7	30209	4316	100***
b22	7	5136	734	6.6**
a3	7	52	8	57***
b3	7	0.2	0	0.97 ^{ns}
Hinge 2	7	3406	487	15***

a1, a2, a22, and a3 correspond to y-intercepts in relation to change in number of days to flowering under the respective photoperiods; b1, b2, b22, and b3, stand for slope coefficients in relation to the changes in time from seeding to first flowering against time from seeding to first transfer. *** and ** indicates significant difference at $P \leq 0.01$ and 0.001 , respectively and ns = not significant.

Photoperiod-Sensitive Inductive Phase

The two photoperiod-sensitive accessions, [CDC Corinne (S1) and S2 (CDC Frontier (S2))] had higher values for the photoperiod-sensitive inductive phase under SD compared to LD. In these accessions, the photoperiod-sensitive inductive phase under LD ranged from 17 to 38 days, and were 49 to 77 under SD. In the accessions with intermediate ICC 15294 (I1), ICC8621 (I2), ILC1867 (I3), and ICC 8855 (I4) response to photoperiod the photoperiod-sensitive inductive phase under LD ranged from 12 to 19 days, and was 25 to 43 under SD.

For ICCV 96029 (N1), the values of the photoperiod-sensitive inductive phase under LD and SD were 0.1 and 0.0, respectively. In FLIP98-142C (N2), another photoperiod-insensitive accession, the range of the photoperiod-sensitive inductive phase under LD and SD was 7 and 15 days, respectively. In the highly photoperiod-sensitive accessions, the photoperiod-insensitive

TABLE 4 | Means comparison of the hinge 1, a1, b1, a2, b2, hinge 2, and a22, b22, a3, and b3 for eight chickpea accessions evaluated in a reciprocal transfer experiment from LD (16 h light) and SD (10 h light) photoperiod conditions over two time replicates.

Accessions	Hinge 1	a1	a2	b1	b2	Hinge 2	a22	a3	b22	b3
CDC Corinne (S1)	20 ^a	98 ^a	123 ^a	−0.3 ^a	−1.6 ^{bd}	42 ^a	127 ^a	60 ^{ab}	−1.7 ^c	0.0 ^a
CDC Frontier (S2)	16 ^{ab}	101 ^a	120 ^a	0.2 ^a	−1.1 ^{bbd}	47 ^a	117 ^a	75 ^a	−1.1 ^{bc}	−0.2 ^a
ICC 15294 (I1)	20 ^a	76 ^b	115 ^{ab}	−0.3 ^a	−2.0 ^d	45 ^a	96 ^b	34 ^c	−1.4 ^c	0.0 ^a
ICC 8621 (I2)	9 ^{ac}	59 ^d	67 ^{dc}	0.0 ^a	−0.9 ^{bbd}	26 ^{bc}	76 ^c	40 ^{bc}	−1.4 ^c	0.0 ^a
ILC 1687 (I3)	10 ^{abc}	68 ^c	80 ^{bc}	0.1 ^a	−1.2 ^{ddb}	36 ^{ab}	82 ^{bc}	36 ^{bc}	−1.3 ^{bc}	0.1 ^a
ICC 8855 (I4)	3 ^{bc}	34 ^f	31 ^e	0.0 ^a	1.2 ^{ab}	19 ^c	11 ^f	72 ^a	3.6 ^a	−0.3 ^a
ICCV 96029 (N1)	0 ^c	28 ^g	28 ^e	0.0 ^a	0.0 ^{ab}	1 ^d	29 ^e	27 ^c	−3.0 ^d	0.0 ^a
FLIP 98-142C (N2)	9 ^{abc}	40 ^e	44 ^{de}	−0.1 ^a	−0.5 ^{bc}	26 ^{bc}	44 ^d	32 ^c	−0.5 ^{bc}	−0.1 ^a

Values with the same letter in the same column are not significantly different; a1, a2, a22, and a3, stand for the y-intercepts (number of days to flowering); b1, b2, b3, b22, correspond to the slope coefficients for the accessions.

TABLE 6 | Hinge regression for the eight chickpea accessions evaluated in the reciprocal transfer from LD to SD.

Experiment	Accessions	y-intercept (days)	slope coefficients (day/day)	R ² (%)	CV (%)
Days to flowering of the accessions transferred from long to SD	CDC Corinne (S1)	97	−0.55	0.9	9
	CDC Frontier (S2)	104	−0.65	0.94	5
	ICC 15294 (I1)	79	−0.8	0.91	11
	ICC 8621 (I2)	59	−0.45	0.81	9
	ILC1687 (I3)	69	−0.65	0.86	10
	ICC 8855 (I4)	34	−1	0.67	8
	ICCV 96029 (N1)	28	0	0.33	2
	FLIP98-142C (N2)	42	−0.4	0.88	5
Days to flowering of the accessions transferred from short to LD	CDC Corinne (S1)	53	0.6	0.92	6
	CDC Frontier (S2)	58	0.6	0.92	6
	ICC 15294 (I1)	31	0.7	0.92	8
	ICC 8621 (I2)	40	0.4	0.92	5
	ILC1687 (I3)	39	0.6	0.9	7
	ICC 8855 (I4)	73	0.95	0.85	11
	ICCV 96029 (N1)	28	0.05	0.64	2
	FLIP98-142C (N2)	29	0.3	0.77	5

The y-intercept and slopes presented here are combined over two time replicates in growth chamber. CV, coefficient of variation.

TABLE 7 | Duration in days of each of the three developmental phases (I_{1L} , I_{1S} , I_{2L} , I_{2S} , and I_{3L} and I_{3S}) \pm SE from seeding to first flower appearance in eight chickpea accessions under LD and SD conditions.

Accessions	I_{1L}	I_{1S}	I_{2L}	I_{2S}	I_{3L}	I_{3S}	f_L	f_S	R ²
CDC Corinne (S1)	19 \pm 3	17 \pm 7	17 \pm 6	49 \pm 14	20 \pm 5	29 \pm 12	56	95	0.99
CDC Frontier (S2)	15 \pm 7	20 \pm 12	38 \pm 9	77 \pm 16	12 \pm 6	4 \pm 10	66	101	0.98
ICC 15294 (I1)	13 \pm 9	18 \pm 15	19 \pm 11	43 \pm 22	1.4 \pm 7	15 \pm 18	35	76	0.97
ICC 8621 (I2)	10 \pm 1	13 \pm 3	8 \pm 2	25 \pm 4	23 \pm 1	23 \pm 2	42	61	0.99
ILC 1687 (I3)	12 \pm 10	12 \pm 5	43 \pm 14	16 \pm 9	19 \pm 10	17 \pm 6	45	74	0.95
ICC 8855 (I4)	9 \pm 5	14 \pm 3	18 \pm 6	32 \pm 5	11 \pm 4	19 \pm 8	38	65	0.97
ICCV 96029 (N1)	22 \pm 0.5	23 \pm 0.3	0.1 \pm 0.9	0	5 \pm 0.9	5 \pm 0	27	28	0.74
FLIP98-142C (N2)	18 \pm 2	20 \pm 3	7 \pm 4	15 \pm 5	5 \pm 2	6 \pm 3	30	41	0.98

The values for the three developmental phases were derived from the experiment conducted in two time replicates and the mean values were used in the model to derive the values for each accession. I_{1L} = photoperiod-insensitive pre-inductive phase under LD, I_{1S} = photoperiod-insensitive pre-inductive phase under SD. I_{2L} = photoperiod-sensitive inductive phase under LD, I_{2S} = photoperiod-sensitive inductive phase under SD. I_{3L} = photoperiod-insensitive post-inductive phase under LD, I_{3S} = photoperiod-insensitive post-inductive phase under SD. f_L = days from seeding to flowering under LD, f_S = days from seedling seeding to flowering under SD. R² = the amount of variation accounted for by the model.

post-inductive phase ranged between 13 and 20 days under LD, and between 4 and 29 days under SD, respectively. In the intermediate accessions, this phase ranged between 5 and 20 days under LD, and between 15 and 23 days under SD. The photoperiod-insensitive accessions had a similar range for the photoperiod-insensitive post-inductive phase of 5–6 days in long as well as in SD.

Photoperiod-Insensitive Post-inductive Phase

In the highly photoperiod-sensitive accessions, the photoperiod-insensitive post-inductive phases were between 13 and 20 days under LD and between 4 and 29 days under SD, respectively. In the intermediate accessions, the values of this phase ranges between 5 and 20 days under LD and between 15 and 23 days under SD. The photoperiod-insensitive accessions had similar range of photoperiod-insensitive post-inductive phases of 5 to 6 days in long as well as SD.

DISCUSSION

When control plants in the respective chambers were compared, the plants under LD flowered earlier than those under SD. Early transfer of plants from either long to SD chambers or vice versa had no effect on the flowering response of the plants. Differences in the number of days to flower between SD and LD control plants were wider compared to the number of days to flower bud initiation. This indicated that day to full flower opening was delayed by short photoperiods after flower bud initiation. Wallace et al. (1993) reported that time to initiation of flower buds could not be used to differentiate the insensitive and sensitive genotypes in soybean.

The slope coefficients of the flowering responses under non-optimal photoperiods provided an estimate of photoperiod sensitivity (Major, 1980). In our study, the absolute values of the slopes for the photoperiod-sensitive and intermediates were

greater than the day-neutral ones. ICCV 96029 (N1) specifically had slopes of 0 and 0.05 for transfers from LD to SD and SD to LD, respectively. Both values were not significantly different from 0, supporting the previous report (Daba et al., 2015) that this accession is day-neutral under a mean temperature of 19°C combined with either 10 or 16 h photoperiod.

The hinge regression function technique was exploited to identify photoperiod-sensitive and -insensitive phases in the chickpea accessions. The hinge technique was very efficient in differentiating between photoperiod-sensitive and -insensitive phases in the photoperiod-sensitive and intermediate accessions. The actual data used to determine the photoperiod sensitivity in most plants seldom resemble the idealized schematic diagram similar to the response of FLIP 98-142C (N2) and CDC Frontier (S2) in our research. Hinge regression functions are applied most importantly in multivariate regression and classifications. In our study, the advantage of the hinge regression function was evident in the day-neutral accession ICC V96029 (N1) for which the first and second hinges were 0 and 1, respectively, indicating the absence of a significant change in the flowering response, and confirming that ICCV 96029 is day-neutral under our experimental conditions.

Chickpea, therefore, has three flowering induction phases: a photoperiod-insensitive pre-inductive phase, a photoperiod-sensitive inductive phase, and photoperiod-insensitive post-inductive phase. An inverse relationship between photoperiod sensitivity phase and photoperiod was identified, i.e., a longer photoperiod-sensitive phase was observed under SD, and a shorter photoperiod-sensitive phase was observed under LD. Variability in the length of the photoperiod-insensitive pre-inductive phase was observed among the photoperiod-sensitive, intermediate, and day-neutral chickpea accessions. A shorter duration of photoperiod-insensitive pre-inductive phase was detected compared to the photoperiod-sensitive phase in intermediate accessions. During the photoperiod-insensitive pre-inductive phase, plants were not responsive to changes in the photoperiod. In many crops, a minimum vegetative period, known as the basic vegetative phase, is required during which there is no response to photoperiod (Vergara and Chang, 1985).

The two high yielding accessions developed and released by the Crop Development Centre, University of Saskatchewan [CDC Corinne (S1) and CDC Frontier (S2)] (Warkentin et al., 2005; Tar'an et al., 2009) had the longest time to flowering, as well as longer duration of photoperiod sensitivity phases under LD and SD. Daba et al. (2015, 2016) reported that ICCV 96029 (N1) and FLIP98-142C (N2) flowered the earliest; ICC 15294 (I1), ICC 8621 (I2), ILC 1867 (I3), and ICC 8855 (I4) flowered intermediate and CDC Frontier (S2) and CDC Corinne (S1) flowered the latest under a combination of temperature and photoperiod in the growth chamber conditions.

Efforts to develop early flowering cultivars adapted to the short growing season of western Canada could exploit ICCV 96029 (N1) and FLIP 98-142C (N2), which have a minimal photoperiod-sensitive phase. This strategy was also recommended by Kumar and Abbo (2001). Photoperiod-insensitivity contributed a significant share for chickpea

adaptation to low latitude during early domestication (Siddique et al., 2003; Rubio et al., 2004). In chickpea, the number of biological days from emergence to flowering should match the latitude locations based on photoperiod sensitivity (Vadez et al., 2013). Early flowering and maturity in photoperiod-insensitive genotypes in bean has helped to attain higher harvest index compared to the photoperiod-sensitive genotypes (Yourstone et al., 1993).

CONCLUSION

The phenology of chickpea accessions from emergence to first flowering can be divided into three phases: (1) a photoperiod-insensitive pre-inductive phase, (2) a photoperiod-sensitive inductive phase, and (3) a photoperiod-insensitive post-inductive phase. The duration of the photoperiod-insensitive pre-inductive phase was shorter than that of the photoperiod-sensitive inductive phase in chickpea. Photoperiod sensitivity commenced on different days after emergence in different accessions. The photoperiod-sensitive inductive phase extended beyond flowering bud initiation and full flower opening to the stage of full flower development. Flower bud initiation and full flower opening appeared to be sensitive to photoperiod at different times after emergence for different chickpea accessions. Time to flower bud initiation as well as time to full flower opening differentiated photoperiod-insensitive and photoperiod-sensitive accessions. In the cool short seasons of Western Canada, chickpea accessions with a shorter duration both the pre-inductive photoperiod-insensitive and photoperiod-sensitive inductive phases are desirable for adaptation. The day-neutral accessions such as ICCV 96029 and FLIP98-142C are used for developing cultivars fit to the tropics, subtropics, and the Mediterranean regions characterized by short growing seasons delimited by increasing temperatures and reduced soil moisture where short crop duration is desired.

AUTHOR CONTRIBUTIONS

KD conducted the experiments, analyzed, and summarized the results. KD, TW, RB, CT, and BT wrote and finalized the manuscript; BT and TW conceived and directed the project.

FUNDING

The authors thank the Saskatchewan Ministry of Agriculture for financial support of this research.

ACKNOWLEDGMENTS

The technical expertise of the Pulse Breeding Staff at the Crop Sciences Field Lab and the phytotron staff is highly appreciated. We also extend our appreciation to Dr. Xinyou Yin (Wageningen University and Research Centre) for providing the code for non-linear model in SAS.

REFERENCES

- Adams, S. R., Munir, M., Valdes, V. M., Langton, F. A., and Jackson, S. D. (2003). Using flowering times and leaf numbers to model the phases of photoperiod sensitivity in *Antirrhinum majus* L. *Ann. Bot.* 92, 689–696. doi: 10.1093/aob/mcg194
- Adams, S. R., Pearson, S., and Hadley, P. (2001). Improving quantitative flowering models through a better understanding of the phases of photoperiod sensitivity. *J. Exp. Bot.* 52, 655–662.
- Breiman, L. (1993). Hinging hyperplanes for regression, classification, and function approximation. *Inform. Theor. IEEE Trans.* 39, 999–1013. doi: 10.1109/18.256506
- Bunting, A. H. (1975). Time, phenology and the yields of crops. *Weather* 30, 312–325. doi: 10.1002/j.1477-8696.1975.tb05259.x
- Collinson, S. T., Ellis, R. H., Summerfield, R. J., and Roberts, E. H. (1992). Durations of the photoperiod-sensitive and photoperiod-insensitive phases of development to flowering in four cultivars of rice (*Oryza sativa* L.). *Ann. Bot.* 70, 339–346.
- Daba, K., Tar'an, B., Bueckert, R., and Warkentin, T. D. (2016). Effect of temperature and photoperiod on time to flowering in Chickpea. *Crop Sci.* 56, 200–208. doi: 10.2135/cropsci2015.07.0445
- Daba, K., Tar'an, B., and Warkentin, T. D. (2015). Flowering response of diverse chickpea (*Cicer arietinum* L.) accessions to photoperiod. *Genet. Res. Crop Evol.* 1–12. doi: 10.1007/s10722-015-0308-5
- Dunnigan, G. M., Hammen, J. L., and Harris, T. R. (1997). A SAS-IML program for implementing two-phase regression analysis of geophysical time series data. *Comput. Geosci.* 23, 763–770. doi: 10.1016/S0098-3004(97)00049-6
- Ellis, R. H., Collinson, S. T., Hudson, D., and Patefield, W. M. (1992). The analysis of reciprocal transfer experiments to estimate the durations of the photoperiod-sensitive and photoperiod-insensitive phases of plant development: an example in soya bean. *Ann. Bot.* 70, 87–92.
- Hadley, P., Roberts, E. H., Summerfield, R. J., and Minchin, F. R. (1984). Effects of temperature and photoperiod on flowering in soya bean [*Glycine max* (L.) Merrill]: a quantitative model. *Ann. Bot.* 53, 669–681.
- Kenesei, T., and Abonyi, J. (2013). Hinging hyperplane based regression tree identified by fuzzy clustering and its application. *Appl. Soft Comput.* 13, 782–792. doi: 10.1016/j.asoc.2012.09.027
- Kiniry, J. R., Ritchie, J. T., Musser, R. L., Flint, E. P., and Iwig, W. C. (1983). The photoperiod sensitive interval in maize. *Agron. J.* 75, 687–690. doi: 10.2134/agronj1983.00021962007500040026x
- Kumar, J., and Abbo, S. (2001). Genetics of flowering time in chickpea and its bearing on productivity in the semi-arid environments. *Adv. Agron.* 72, 107–138. doi: 10.1016/S0065-2113(01)72012-3
- Lejeune-Hénaut, I., Bourion, V., Etévé, G., Cunot, E., Delhay, K., and Desmyter, C. (1999). Floral initiation in field-grown forage peas is delayed to a greater extent by short photoperiods, than in other types of European varieties. *Euphytica* 109, 201–211. doi: 10.1023/A:1003727324475
- Levy, Y. Y., and Dean, C. (1998). The transition to flowering. *Plant Cell* 10, 1973–1990. doi: 10.1105/tpc.10.12.1973
- Major, D. J. (1980). Photoperiod response characteristics controlling flowering of nine crop species. *Can. J. Plant Sci.* 60, 777–784. doi: 10.4141/cjps80-115
- Putterill, J., Laurie, R., and Macknight, R. (2004). It's time to flower: the genetic control of flowering time. *Bioessays* 26, 363–373. doi: 10.1002/bies.20021
- Ritchie, J. T. (1991). "Wheat phasic development," in *Modeling Plant and Soil System, Agronomy Monograph No. 31*, eds J. Hanks and J. T. Ritchie (Madison, WI: American Society of Agronomy), 31–54.
- Ritchie, J. T., Singh, U., Godwin, D. C., and Bowen, W. T. (1998). "Cereal growth, development and yield," in *Understanding Options for Agricultural Production*, eds G. Y. Tsuji, G. Hoogenboom, and P. K. Thornton (Dordrecht: Kluwer Academic Publishers), 79–98.
- Roberts, E. H., Hadley, P., and Summerfield, R. J. (1985). Effects of temperature and photoperiod on flowering in chickpeas (*Cicer arietinum* L.). *Ann. Bot.* 55, 881–892.
- Rubio, J., Flores, F., Moreno, M., Cubero, J., and Gil, J. (2004). Effects of the erect/bushy habit, single/double pod and late/early flowering genes on yield and seed size and their stability in chickpea. *Field Crops Res.* 90, 255–262. doi: 10.1016/j.fcr.2004.03.005
- SAS Institute Inc. (2009). *SAS-Statistical Analysis Software for Windows Version 9.3*. Cary, NC: SAS Inst. Inc.
- Shaykewich, C. F. (1995). An appraisal of cereal crop phenology modelling. *Can. J. Plant Sci.* 75, 329–341. doi: 10.4141/cjps95-057
- Siddique, K. H. M., Loss, S. P., Thomson, B. D., and Saxena, N. P. (2003). "Cool season grain legumes in dryland Mediterranean environments of Western Australia: significance of early flowering," in *Management of Agricultural Drought: Agronomic and Genetic Options*, ed. N. P. Saxena (Enfield, NH: Science Publishers), 151–162.
- Slafer, G. A., and Rawson, H. M. (1994). Sensitivity of wheat phasic development to major environmental factors: a re-examination of some assumptions made by physiologists and modellers. *Funct. Plant Biol.* 21, 393–426. doi: 10.1071/pp9940393
- Slafer, G. A., and Rawson, H. M. (1995). Development in wheat as affected by timing and length of exposure to long photoperiod. *J. Exp. Bot.* 46, 1877–1886. doi: 10.1093/jxb/46.12.1877
- Slafer, G. A., and Rawson, H. M. (1997). Phyllochron in wheat as affected by photoperiod under two temperature regimes. *Funct. Plant Biol.* 24, 151–158. doi: 10.1071/pp96021
- Soltani, A., Torabi, B., Zeinali, E., and Sarparast, R. (2004). Response of chickpea to photoperiod as a qualitative long-day plant. *Asian J. Plant Sci.* 3, 705–708. doi: 10.3923/ajps.2004.705.708
- Summerfield, R. J., Ellis, R. H., and Craufurd, P. Q. (1996). Phenological adaptation to cropping environment. From evaluation descriptors of times to flowering to the genetic characterization of flowering responses to photoperiod and temperature. *Euphytica* 92, 281–286. doi: 10.1007/BF00022856
- Tar'an, B., Warkentin, T. D., Banniza, S., and Vandenberg, A. (2009). CDC Corinne desi chickpea. *Can. J. Plant Sci.* 89, 515–516. doi: 10.4141/CJPS08206
- Vadez, V., Soltani, A., and Sinclair, T. R. (2013). Crop simulation analysis of phenological adaptation of chickpea to different latitudes of India. *Field Crops Res.* 146, 1–9. doi: 10.1016/j.fcr.2013.03.005
- Vergara, B. S., and Chang, T. T. (1985). *The Flowering Response of the Rice Plant to Photoperiod. A Review of the Literature*, 4th Edn. Los Banos, CA: The International Rice Research Institute.
- Wallace, D., Zobel, R., and Yourstone, K. (1993). A whole-system reconsideration of paradigms about photoperiod and temperature control of crop yield. *Theor. Appl. Genet.* 86, 17–26. doi: 10.1007/BF00223804
- Wang, Z., Acock, M. C., and Acock, B. A. (1997). Photoperiod sensitivity during flower development of Opium Poppy (*Papaver somniferum* L.). *Ann. Bot.* 79, 129–132. doi: 10.1006/anbo.1996.0319
- Warkentin, T. D., Banniza, S., and Vandenberg, A. (2005). CDC Frontier kabuli chickpea. *Can. J. Plant Sci.* 85, 909–910. doi: 10.4141/P04-185
- Warkentin, T. D., Vandenberg, A., Banniza, S., Tar'an, B., Tullu, A., Lulsdorf, M., et al. (2003). "Breeding chickpea for improved *Ascochyta* blight resistance and early maturity in western Canada," in *Proceedings of International Chickpea Conference; 2003 20-22 January*, eds R. N. Sharma, M. Yasin, S. L. Swami, M. A. Khan and A. J. William (Raipur: Indira Gandhi Agricultural University), 1–4.
- Warner, R. M. (2009). Determination of photoperiod-sensitive stages of development of the short-day plant *Celosia*. *HortScience* 44, 328–333.
- Yin, X. (2008). Analysis of reciprocal-transfer experiments to estimate the length of phases having different responses to temperature. *Ann. Bot.* 101, 603–611. doi: 10.1093/aob/mcm319
- Yin, X., Kropff, M. J., Horie, T., Nakagawa, H., Centeno, H. G. S., Zhu, D., et al. (1997). A model for photothermal responses of flowering in SnRice SnI. Model description and parameterization. *Field Crops Res.* 51, 189–200. doi: 10.1016/S0378-4290(96)03456-9
- Yourstone, K. S., Masaya, P. N., and Zobel, R. W. (1993). Photoperiod gene control over partitioning between reproductive and vegetative growth. *Theor. Appl. Genet.* 86, 6–16.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Daba, Warkentin, Bueckert, Todd and Tar'an. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Flowering and Growth Responses of Cultivated Lentil and Wild *Lens* Germplasm toward the Differences in Red to Far-Red Ratio and Photosynthetically Active Radiation

Hai Y. Yuan, Shyamali Saha, Albert Vandenberg and Kirstin E. Bett*

Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:

Nicolas Rispaïl,
Consejo Superior de Investigaciones
Científicas, Spain

Reviewed by:

Dharmendra Singh,
Indian Council of Agricultural
Research, India
Ane Victoria Vollsnes,
University of Oslo, Norway

*Correspondence:

Kirstin E. Bett
k.bett@usask.ca

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 19 December 2016

Accepted: 07 March 2017

Published: 21 March 2017

Citation:

Yuan HY, Saha S, Vandenberg A and
Bett KE (2017) Flowering and Growth
Responses of Cultivated Lentil
and Wild *Lens* Germplasm toward
the Differences in Red to Far-Red
Ratio and Photosynthetically Active
Radiation. *Front. Plant Sci.* 8:386.
doi: 10.3389/fpls.2017.00386

Understanding environmental responses of pulse crop species and their wild relatives will play an important role in developing genetic strategies for crop improvement in response to changes in climate. This study examined how cultivated lentil and wild *Lens* germplasm responded to different light environments, specifically differences in red/far-red ratio (R/FR) and photosynthetically active radiation (PAR). Three genotypes of each the seven *Lens* species were grown in environmentally controlled growth chambers equipped to provide light treatments consisting of different R/FR ratios and PAR values. Our results showed that overall, days to flower of *Lens* genotypes were mainly influenced by the R/FR induced light quality change but not by the PAR related light intensity change. The cultivated lentil (*L. culinaris*) showed consistent, accelerated flowering in response to the low R/FR light environment together with three wild lentil genotypes (*L. orientalis* IG 72611, *L. tomentosus* IG 72830, and *L. ervoides* IG 72815) while most wild lentil genotypes had reduced responses and flowering time was not significantly affected. The longest shoot length, longest internode length, and largest leaflet area were observed under the low R/FR low PAR environment for both cultivated and wild lentils. The distinctly different responses between flowering time and elongation under low R/FR conditions among wild *Lens* genotypes suggests discrete pathways controlling flowering and elongation, which are both components of shade avoidance responses. The yield and above-ground biomass of *Lens* genotypes were the highest under high R/FR high PAR conditions, intermediate under low R/FR low PAR conditions, and lowest under high R/FR low PAR light conditions. Three *L. lamottei* genotypes (IG 110809, IG 110810, and IG 110813) and one *L. ervoides* genotype (IG 72646) were less sensitive in their time to flower responses while maintaining similar yield, biomass, and harvest index across all three light environments; these are indications of better adaptability toward changes in light environment.

Keywords: lentil, red/far-red ratio, photosynthetically active radiation, photomorphogenesis, photosynthesis

INTRODUCTION

Light is essential for plant growth and development. It is a source of energy for photosynthesis and provides information for regulation of growth and development (Smith, 1982). Light is necessary for the adaptation of plants to specific environments. Information based on light quality induces a collective photomorphogenetic response (Quail, 2002), which is controlled by several photoreceptors including blue-light absorbing phototropin and cryptochrome and the red and far-red light absorbing phytochrome family (Rockwell et al., 2006; Möglicher et al., 2010). Among these photoreceptors, phytochromes play a key role in photomorphogenesis and control about 10% of the plant transcriptome (Quail, 2002). Phytochromes exist in two forms, the red-light absorbing Pr, which absorbs maximally at 660 nm and is generally considered to be biologically inactive, while the far-red light absorbing Pfr, which absorbs maximally at 730 nm and is biologically active (Franklin and Whitelam, 2005). Absorption of light by either Pr or Pfr results in phototransformation between these two forms, which drives the on/off switching of the successive signaling pathway (Han et al., 2007).

Green plants selectively absorb blue and red wavelengths through chlorophyll and carotenoid photosynthetic pigments. The radiation reflected by green leaves is relatively enriched in the far-red spectrum (Smith and Whitelam, 1997). A resulting decrease in red/far-red (R/FR) ratio in the surrounding environment is sensed by the phytochromes, thereby signaling the presence of neighboring plants as potential competition. This leads to enhanced stem elongation and accelerated transition to flowering, which are features of the shade avoidance response (SAR) (Whitelam and Devlin, 1997). A reduction in light intensity usually triggers a similar SAR (Ballare et al., 1991). Under natural conditions, the SAR allows plants to compete with neighboring vegetation for limited resources (Schmitt et al., 2003). For crop species, however, SAR could lead to decreased yield if plants spend resources on vegetative growth at the expense of reproductive development (Board, 2000; Kebrom and Brutnell, 2007; Casal, 2013). Stem elongation often leads to lodging, as observed by Gong et al. (2015) in soybean (*Glycine max*), and genotypes displaying reduced SAR are suggested to have improved performance at higher plant densities (Kebrom and Brutnell, 2007; Gong et al., 2015).

A close relationship between light intensity and plant growth and yield has been described for many plant species, including legumes (Bethlenfalvay and Phillips, 1977; Baligar et al., 2008; Mielke and Schaffer, 2010). The literature contains a few reports on flowering and growth of legume species in response to light quality changes (Board, 2000; Heraut-Bron et al., 2000; Gong et al., 2015), but little published information exists on the interactive effects of light intensity and light quality.

Lentil (*Lens culinaris* Medik.) is a quantitative long-day species (Summerfield et al., 1985). Lentil plays a significant role in supporting environmentally sustainable agriculture through its nitrogen fixation ability, and is internationally recognized as part of the solution to global food and nutritional insecurity. The genus *Lens* has seven species that originated around the

Mediterranean and into the Middle East and the domesticated lentil is *L. culinaris* (Cubero et al., 2009). The other six wild species within the genus *Lens* are *L. culinaris*, *L. orientalis*, *L. tomentosus*, *L. odemensis*, *L. lamottei*, *L. ervoides*, and *L. nigricans* (Wong et al., 2015). Resistance to biotic or abiotic stresses that are lacking in cultivated lentil have been identified in wild relatives (Tullu et al., 2006, 2010; Fiala et al., 2009; Podder et al., 2012; Vail et al., 2012) and this genetic resource is considered to have great value for future genetic improvement.

Growth and development of cultivated lentil was significantly affected when standard lighting (fluorescent and incandescent bulbs) was replaced with high efficiency and high output fluorescent bulbs in our controlled environment growth chambers. Flowering of cultivated lentil was delayed by up to 30 days (Mobini et al., 2012, 2016; Yuan et al., 2015) but some wild germplasm seemed unaffected. The new light system affected both light quality (increased R/FR) and light intensity [increased photosynthetically active radiation (PAR)]. As a first step to gaining an understanding of the genetic basis of light responses in genus *Lens*, we used a collection of wild and cultivated *Lens* germplasm to characterize the flowering and growth responses to changes in R/FR and PAR. The hypothesis was that flowering and growth development in *Lens* germplasm could be different due to the differences in R/FR and PAR and if so, these genotypes would represent key genetic resources for developing lentil cultivars with better adaptation to variable light environments.

MATERIALS AND METHODS

Plant Materials

Three genotypes of each of the seven *Lens* species were selected from germplasm currently in use in the breeding program at the Crop Development Centre (CDC), University of Saskatchewan (USASK) (Table 1). Prior to planting, seeds of all genotypes were stored at -20°C for 1 week and then scarified to improve imbibition and germination. Square 10-cm pots were filled with growth medium consisting of 50% Sunshine[®] Mix #3 and 50% Sunshine[®] Mix #4 (Sun Gro Horticulture Canada, Ltd, Seba Beach, AB, Canada¹). Two seeds were sown in each pot and thinned to a single plant after emergence. Each repeat had four pots (replicates) per genotype and the experiment was repeated once. All plants were grown and maintained in controlled growth cabinets set at $22^{\circ}\text{C}/16\text{ h day}$ and $16^{\circ}\text{C}/8\text{ h night}$ for both repeats.

Light Environment and Growing Conditions

Two Conviron GR48 walk-in plant growth chambers and one Conviron PGV36 walk-in plant growth chamber were used in the experiment. The light properties of the three growing environments are described in Table 2. Spectral distribution and characteristics of the light treatments used in the experiment are shown in Figure 1. The high R/FR was the natural ratio of the renovated light system in the growth chambers fitted with T5 841 High Output Fluorescence bulbs (Philips, Andover, MA, USA).

¹<http://www.sungro.com/professional-products>

The low R/FR environment was fitted with evenly spaced infrared bulbs (Model FLR48T5NIR-HO, DN Lighting, Co. Ltd, Hiratsuka City, Japan) to replace the T5 841 fluorescent bulbs in the light bank. The high PAR condition was the natural light intensity of the renovated light system while the low PAR condition was reached by adjusting the height of the light bank. The spectral photon flux and PAR of each light treatment was measured using a spectroradiometer (Apogee Instruments, Model PS-300, Logan, UT, USA). The R and FR values were calculated using the spectral photon flux at 650–670 and 720–740 nm, respectively (Smith, 1982).

Data Collection and Analysis

Lentil development stages were based on Erskine et al. (1990) with minor modifications to suit the wild lentil species. Days to flower were calculated based on days from emergence to R1 (one open flower at any node), and the node number of the first open flower was recorded. Leaflet area was measured using the leaflets produced at the R1 flowering node using a flatbed scanner (Epson Perfection V700 Photo Scanner, Epson Canada, Markham, ON, Canada) and WinFOLIA™ software (Regent Instruments, Inc., Québec City, QC, Canada). Lentil plants produce one leaf per node, but have 5–15 leaflets per leaf depending on species and genotype. Five fully expanded

leaflets beginning from the base of the leaf at the R1 flowering node from four plants (replicates) were removed from the plants and kept in Petri-dishes on top of ice and leaflet area was immediately measured. Shoot length and node number were recorded when the plants reached R5 (one mature pod at any node), at which point mesh bags were placed around the plants to collect seeds at maturity. Average internode length was calculated using node number at R5 divided by shoot length. The time from sowing to harvest was kept at 110 days, which corresponds to the regular growing period for field grown lentil in Western Canada. Above-ground plant materials were harvested separately for each plant and dried in the dryer. Above-ground biomass was recorded when plant materials were totally dried. Yield (seed production per plant) was recorded after threshing. Harvest index was calculated using the yield divided by the above-ground biomass. Data were collected from two repeats (four reps each), except for leaflet analysis, above-ground biomass, and harvest index which were recorded for the second repeat only. The General Mixed Model Procedure (PROC MIXED) in SAS (SAS Institute, Inc., Cary, NC, USA) was used to determine the effects of genotype, light treatment, and their interaction on observed traits. Paired comparisons were performed using Tukey's honestly significant test, with p -values ≤ 0.05 considered significant.

TABLE 1 | Lentil species and genotypes used in light experiments.¹

Species	Genotypes	Country of origin	Geolocation (lat, long)	Elevation (m)	Seed source
<i>L. culinaris</i>	Eston	Canada	52.8, –106.4	481	CDC, USASK
	CDC Greenstar	Canada	52.8, –106.4	481	
	CDC QG-2	Canada	52.8, –106.4	481	
<i>L. orientalis</i>	IG 72529	Turkey	38.7, 39.2	1132	ICARDA
	IG 72611	Turkey	38.7, 42.2	2117	ICARDA
	BGE 016880	Israel	N/A	N/A	Universidad de León ²
<i>L. tomentosus</i>	IG 72614	Turkey	37.9, 40.3	662	ICARDA
	IG 72805	Turkey	37.8, 39.8	1150	ICARDA
	IG 72830	Turkey	37.7, 39.2	1250	ICARDA
<i>L. odemensis</i>	IG 72623	Turkey	37.4, 41.0	1001	ICARDA
	IG 72639	Syria	34.8, 36.4	950	ICARDA
	IG 72760	Syria	32.7, 36.6	1100	ICARDA
<i>L. lamottei</i>	IG 110809	Spain	36.8, –5.7	150	ICARDA
	IG 110810	Spain	36.8, –5.4	800	ICARDA
	IG 110813	Spain	37.4, –4.3	660	ICARDA
<i>L. ervoides</i>	IG 72646	Syria	35.8, 36.0	570	ICARDA
	IG 72815	Turkey	37.6, 36.5	860	ICARDA
	L01-827A	Jordan ³	30.8, 35.6 ³	1250 ³	CDC, USASK ⁴
<i>L. nigricans</i>	IG 116024	Turkey	37.8, 29.1	560	ICARDA
	IG 136640	Italy	45.1, 7.1	1464	ICARDA
	IG 72555	Spain	37.3, –2.5	1314	ICARDA

¹Information in this table was extracted from Genesys (<https://www.genesys-pgr.org/welcome>). ²Fratini et al. (2007). ³Geolocation and elevation are from *L. orientalis* IG 72847. ⁴Fiala et al. (2009); L01-827A is a single plant selection from a packet of *L. orientalis* IG 72847 from ICARDA.

TABLE 2 | Light properties of the three growing environments used in the current experiment.

Treatments	Low R:FR Low PAR	High R:FR Low PAR	High R:FR High PAR
Red _{650–670} /Far-red _{720–740} ratio	1.6	8.1	8.1
Photosynthetically active radiation (PAR, $\mu\text{mole/m}^2\cdot\text{s}$)	420	420	680

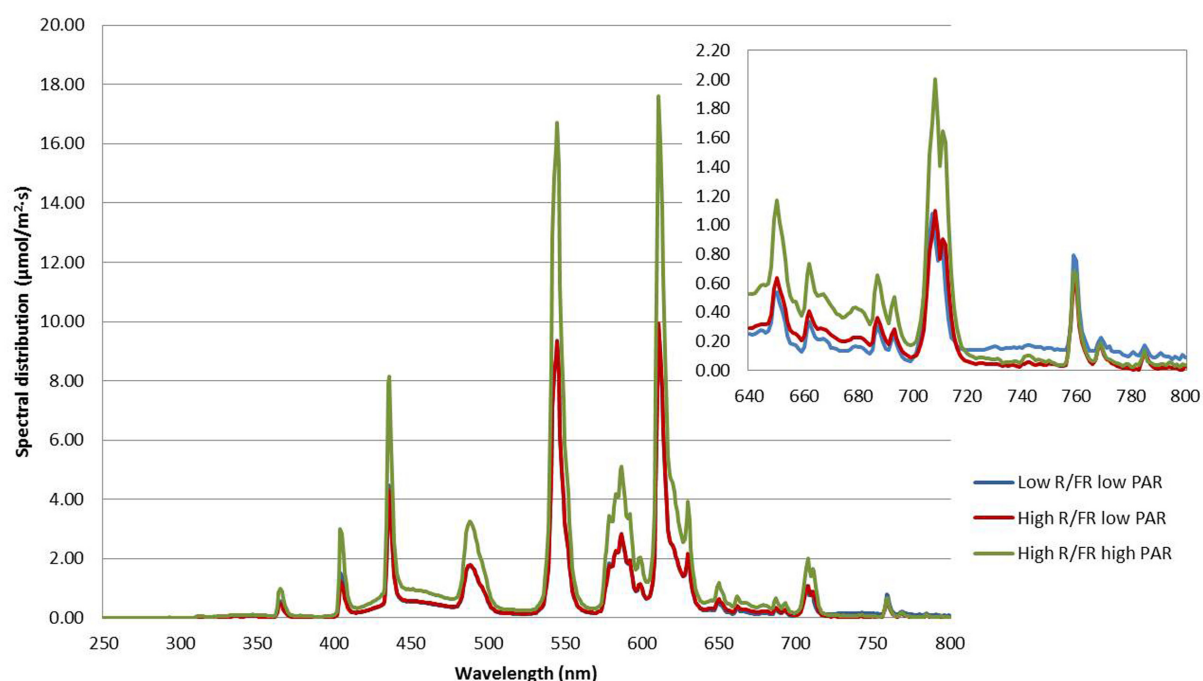


FIGURE 1 | Spectral distribution and characteristics of the light treatments used to compare the effect of light quality and quantity on lentil growth and development. The upper inset shows an enlarged version of the 640–800 nm region.

RESULTS

Overall Growth and Development in *Lens* Genotypes Subjected to Different Light Environments

Growth and development of each genotype varied among light environments. The light environment did not alter the time to emergence but affected the development time from VE (emergence) to R1 and from R1 to R5 in some genotypes (figure shown in Supplementary Materials). Typically, high R/FR delayed the development from VE to R1, while low PAR delayed the development from R1 to R5, if there was an effect. One extreme case was *L. ervoides* IG 72815, for which a mean 35-day difference from the VE to R5 stage was found between

the low R/FR low PAR environment and the high R/FR low PAR environment. Of the three *L. nigricans* genotypes, two did not reach R1 under any light environment, while *L. nigricans* IG 72555 flowered after 94 days in the low R/FR low PAR environment. Therefore, these three genotypes were not included in further data analyses.

Flowering in *Lens* Genotypes to Changes in Light Environment

Days to flower (DTF) were significantly influenced by the genotype, the light environment, and the interaction between them (Table 3). Overall, significant differences were noted between low R/FR and high R/FR environments, while no difference in DTF was detected between low PAR and high

TABLE 3 | Results of the two-way ANOVA for growth characteristics of *Lens* genotypes grown under three different light regimes (treatment).

Growth characteristic	Genotype		Treatment		Genotype × Treatment	
	F	Pr > F	F	Pr > F	F	Pr > F
Days to flower	65.32	<0.0001	81.11	<0.0001	2.53	<0.0001
Shoot length	19.53	<0.0001	291.60	<0.0001	2.07	0.0006
Internode length	12.29	<0.0001	271.95	<0.0001	1.21	0.2033
Leaflet area	31.47	<0.0001	73.39	<0.0001	5.97	<0.0001
Yield	7.30	<0.0001	73.72	<0.0001	1.96	0.0021
Biomass	7.64	<0.0001	85.46	<0.0001	1.60	0.0297
Harvest index	22.75	<0.0001	74.47	<0.0001	3.07	<0.0001

Non-significant factors are bolded.

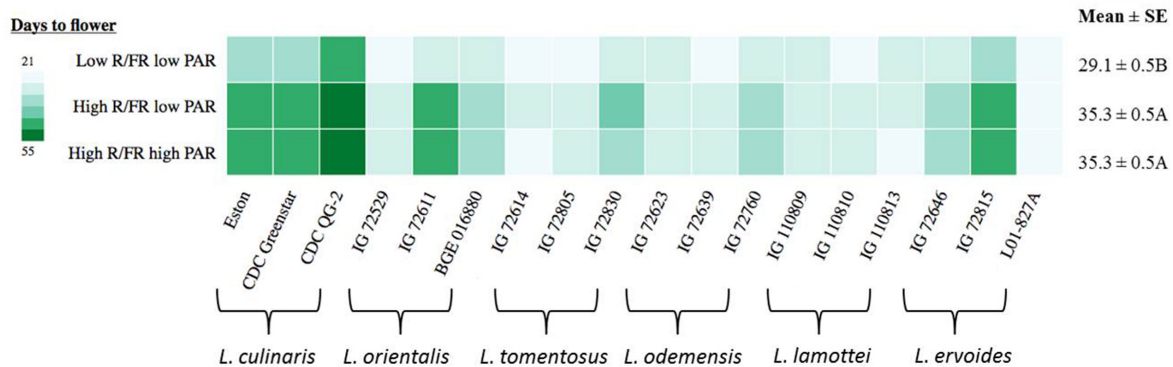


FIGURE 2 | Days to flower for three genotypes of each of six *Lens* species grown in light environments differing in red/far-red ratio (R/FR) and PAR. Mean ± SE is from two repeats (eight replicates) and represents all *Lens* genotypes grown in the specific light environment. Different letters represent significant differences among treatments (Tukey's test at $p \leq 0.05$). Heat map is generated using the in-house developed computer program and the monochrome color coding system was used with the color from lighter to darker represented the values from lower to higher.

PAR when grown under high R/FR conditions (Figure 2). Days to flower in cultivated lentil genotypes (*L. culinaris*: Eston, CDC Greenstar, and CDC QG-2) were all significantly ($p \leq 0.05$) affected by the R/FR induced light quality change but not the PAR related light intensity change. Days to flower in wild lentil species was affected less by changes in either light quality or light intensity; the exceptions were *L. orientalis* IG 72611, *L. tomentosus* IG 72830, and *L. ervoides* IG 72815 for which flowering was significantly ($p \leq 0.05$) delayed under the high R/FR environment like the cultivated lentil.

Shoot Length, Internode Length, and Leaflet Area of *Lens* Genotypes to Changes in Light Environment

Both shoot length and internode length of *Lens* genotypes were significantly influenced by genotype, the light environment, and the interaction between them, except for the genotype \times treatment interaction on internode length (Table 3). Significant differences were noted among species for shoot length and internode length and these differences were more distinct under low R/FR conditions (Figure 3). Low R/FR light resulted in significantly longer shoot and internode length compared to the high R/FR environment ($p \leq 0.05$, Figure 3). Under high R/FR condition, high PAR resulted in significant longer shoots and internodes compared to low PAR ($p \leq 0.05$, Figure 3). Overall, wild lentils had similar treatment responses compared to cultivated lentil for these characteristics.

Leaflet area at the R1 node had the same trend as shoot and internode length, being significantly influenced by genotype, the light environment, and the interaction between them (Table 3). The largest leaflet area resulted from growth in the low R/FR low PAR environment ($p \leq 0.05$; Figure 3), except for *L. culinaris* CDC Greenstar for which a high R/FR environment with a high PAR resulted in larger leaflet area compared to a low R/FR with low PAR.

Yield, above Ground Biomass, and Harvest Index of *Lens* Genotypes to Changes in Light Environment

Both yield and above-ground biomass of *Lens* genotypes were significantly influenced by the different light treatments and both traits showed similar trends across the different treatments (Table 3 and Figure 4). Overall, *Lens* genotypes had the highest yield and above-ground biomass under the high R/FR high PAR condition, while the low R/FR low PAR was intermediate and the high R/FR low PAR had the lowest ($p \leq 0.05$, Figure 4). However, the yields of some wild lentil genotypes, including three *L. lamottei* genotypes (IG 110809, IG 110810, and IG 110813) and *L. ervoides* IG 72646, were not significantly different across all light treatments ($p \leq 0.05$). The yields of genotypes *L. orientalis* IG 72611 and *L. ervoides* IG 72815 under the high R/FR condition were not representative of the true yields because the experiments were stopped at 110 days after sowing. The flowering time of these two genotypes were significantly delayed under the high R/FR environment, which therefore delayed both pod setting and maturity. For *L. odemensis* IG 72760, late emergence combined with a low PAR reduced seed yield.

The harvest index of *Lens* genotypes was significantly influenced by genotype, light environment, and the interaction between them (Table 3). Overall, the harvest indices of *Lens* genotypes were not significantly different ($p \leq 0.05$) under low R/FR low PAR and high R/FR high PAR conditions, whereas the high R/FR low PAR condition resulted in a significant decrease ($p \leq 0.05$, Figure 4). However, harvest indices of all three *L. lamottei* genotypes (IG 110809, IG 110810, and IG 110813), *L. ervoides* IG 72646, and *L. tomentosus* IG 72805 were not significantly different across all three light treatments ($p \leq 0.05$). The harvest indices of *L. orientalis* IG 72611 and *L. ervoides* IG 72815 under the high R/FR condition were not representative of the true harvest indices because the experiments were stopped at 110 days after sowing and the delayed flowering of these two genotypes affected the yields as mentioned previously, therefore affected the harvest indices. The harvest index of *L. odemensis* IG

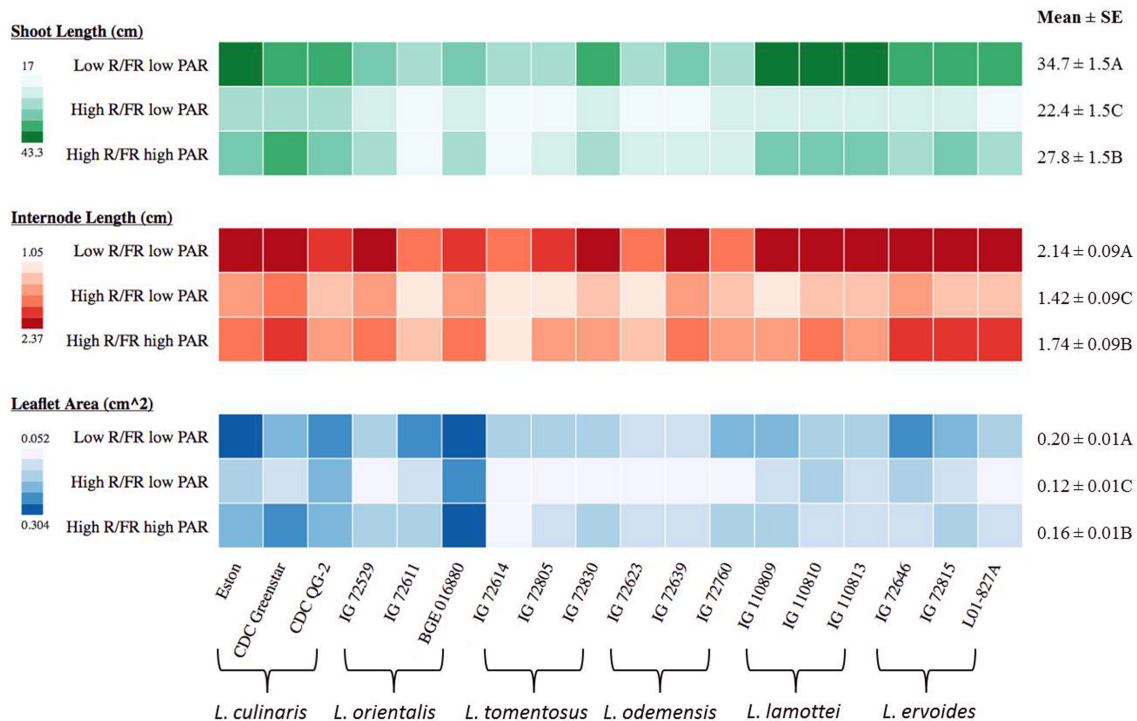


FIGURE 3 | Shoot length, internode length, and leaflet area for three genotypes of each of six *Lens* species grown in light environments differing in red/far-red ratio (R/FR) and photosynthetically active radiation (PAR). Mean ± SE is from two repeats (eight replicates) for shoot length and internode length while one repeat (five replicates) for leaflet area and represents all *Lens* genotypes grown in the specific light environment. Different letters represent significant differences among treatments (Tukey's test at $p \leq 0.05$). Heat map is generated using the in-house developed computer program and the monochrome color coding system was used for each trait with the color from lighter to darker represented the values from lower to higher.

72760 was also not representative of the true harvest index due to affected yield from late emergence and low PAR as mentioned in previous yield section.

DISCUSSION

Flowering Initiation of *Lens* Genotypes is Mainly Influenced by R/FR Related Light Quality Change

Plants depend on the acquisition of light energy for their survival, and competition for light is a characteristic of plant communities. Responses to changes in light quality and intensity enables plants to adapt and optimize their subsequent growth and development. A natural light environment under a canopy has a low R/FR ratio since plants absorb most of the visible light (from 400 to 700 nm) but reflect most of the FR light (Smith, 1982, 1994). A low R/FR ratio reflected from the surrounding vegetation may create a signal to plants of potential competition for light and, therefore, they initiate escape or SARs (Ballare et al., 1990). If the reduced R/FR ratio signal persists and the plant is unable to overcome competing vegetation by growth extension, flowering is accelerated, thereby promoting seed set and enhancing the probability of reproductive success (Halliday et al., 1994; Smith and Whitelam, 1997).

The current study tested seven species of *Lens* and demonstrated that only the cultivated lentil (*L. culinaris*) showed consistent responses to the low R/FR light treatment and this low R/FR light quality promoted early flowering. Most wild lentil genotypes exhibited reduced responses toward the light quality changes and flowering times were not significantly affected. Three wild lentil genotypes (*L. orientalis* IG 72611, *L. tomentosus* IG 72830, and *L. ervoides* IG 72815) had similar flowering responses to the cultivated lentil. However, a genotyping-by-sequencing (GBS) study of the genus *Lens* (Wong et al., 2015) concluded that these three lines fall outside the cluster of other members of those species and may be natural hybrids. This hints at the possibility of transferring the genes controlling the response from one species to another.

A general model for red and far-red light absorbing phytochrome action is that phytochromes perceive light, enter the nucleus, and then interact with transcriptional regulators to regulate gene transcription (Chen et al., 2004; Lorrain et al., 2006; Han et al., 2007). Five members of the phytochrome family were discovered in *Arabidopsis*, named phy A to phy E (Sharrock and Quail, 1989; Clack et al., 1994), which have differential photosensory and physiological roles in controlling plant growth and development (Smith and Whitelam, 1990; Whitelam and Devlin, 1997; Franklin and Quail, 2010). Some of the phytochrome functions elucidated through analysis of

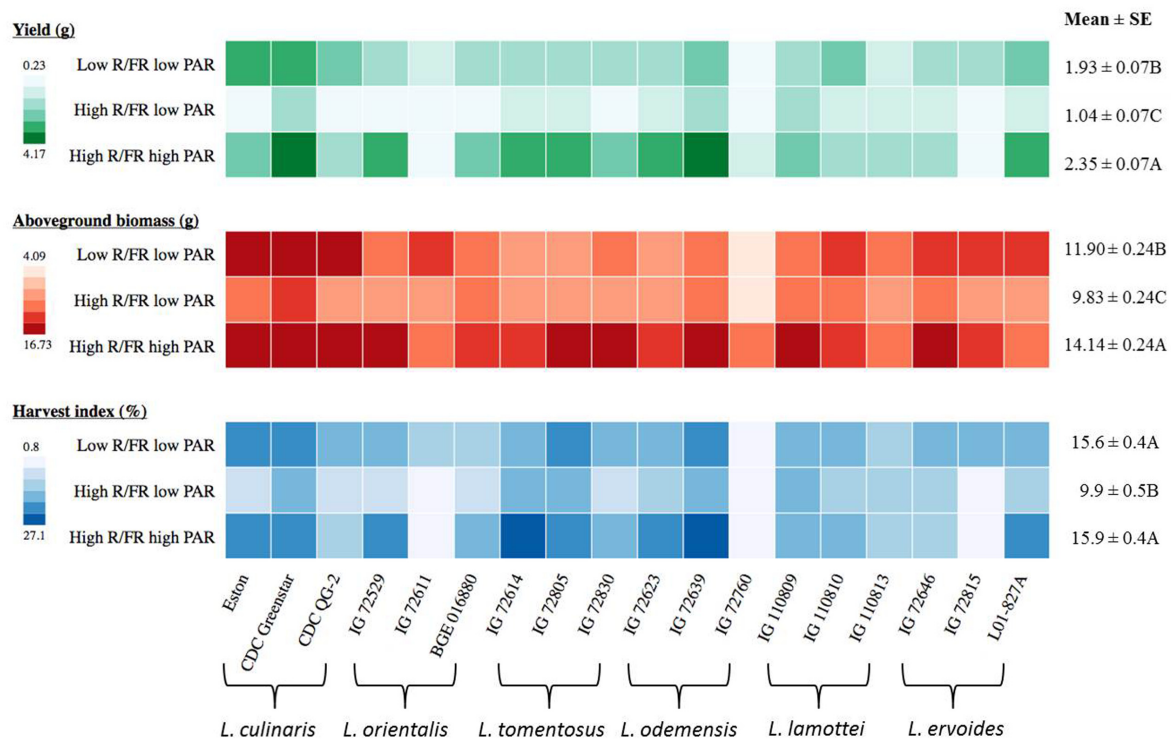


FIGURE 4 | Yield, above-ground biomass, and harvest index for three genotypes of each of six *Lens* species grown in light environments differing in red/far-red ratio (R/FR) and PAR. Mean \pm SE is from two repeats (eight replicates) for yield and one repeat (four replicates) for above-ground biomass and harvest index and it represents all *Lens* genotypes grown in the specific light environment. Different letters represent significant differences among treatments (Tukey's test at $p \leq 0.05$). Heat map is generated using the in-house developed computer program and the monochrome color coding system was used for each trait with the color from lighter to darker represented the values from lower to higher.

Arabidopsis mutants demonstrated that the suppression of flowering under a high R/FR ratio is mediated predominantly by phy B, with redundancy roles for phy D and phy E (Whitelam and Smith, 1991; Halliday et al., 1994; Devlin et al., 1999; Franklin et al., 2003). Genetically distinct signaling pathway segments among the phytochrome family members were also identified (Li et al., 2011). Various studies report that phytochrome genes and flowering genes are well-conserved between *Arabidopsis* and legumes at the level of sequence and physiological function (Hecht et al., 2005; Ueoka-Nakanishi et al., 2011; Wu et al., 2011). Therefore, we suspect that differences in genes of the red and far-red light absorbing phytochrome family and its signaling pathway may also play a direct and important role with respect to the different flowering responses within *Lens* genotypes. Through the domestication process, the variant(s) that make a plant more sensitive to R/FR may have been retained.

The spectral distribution in terms of R/FR ratio of daylight is broadly constant at a specific latitude but varies considerably with geographical location (Smith, 1982). Maloof et al. (2001) reported a correlation between light response and latitude of origin regarding the hypocotyl lengths in 141 *Arabidopsis thaliana* accessions, but no connection regarding the longitude. However, this may not be the case in current study where *Lens nigricans* was the only wild lentil species that showed a large response to change in light environment (no sign of flowering even 110 days

after sowing) yet the latitude of origin for *L. nigricans* is well within the range of other *Lens* species. Previous experience (Saha, unpublished) and results from the current study suggest that favorable flowering conditions for *L. nigricans* may be the low R/FR high PAR condition, which was not assessed here due to limitations of the available lighting systems. A quaternary gene pool placement proposed by Wong et al. (2015) might explain the distinctly different responses of the *L. nigricans* group to the different light quality treatments compared to the other *Lens* species. This species might also have evolved to feature an extended juvenile phase.

Vegetative Growth in *Lens* Genotypes was Mainly Affected by Light Quality with an Interaction of Light Intensity and Controlled by Discrete Pathways from Flowering

The strategy of low R/FR induced shade avoidance used by many plants is to promote growth extension in an attempt to harvest more available light (Runkle and Heins, 2001) and, therefore, the most dramatic SAR is the stimulation of elongation (Devlin et al., 1996; Morelli and Ruberti, 2000). This was clearly shown in our current study. Shoot length and internode length were the longest under the low R/FR environment in all *Lens* genotypes.

Reductions in light intensity usually trigger SARs similar to those under a low R/FR condition (Ballare et al., 1991). In our study, however, we found that high PAR stimulated shoot and internode elongation in *Lens* genotypes in comparison to low PAR under the high R/FR environment, which might be due to the high assimilation rate and high source/sink ratio that occur in the high PAR environment. Evers et al. (2011) observed a similar result, where a high PAR condition promoted branch growth in *Arabidopsis*. Overall, our results show that a low R/FR environment promotes shoot and internode elongation in *Lens* genotypes and high PAR also contributes to elongation. To expand or elongate an organ, as found in a SAR, plants must have a coupling mechanism to process cell division, cell elongation, and cell differentiation. The combined action of plant hormones including gibberellin and auxin play an important role in coordinating the response (Morelli and Ruberti, 2000; Franklin, 2008). The distinctly different responses in flowering time and elongation under low R/FR conditions in the wild lentil species but not in the cultivated lentil suggest discrete pathways control flowering and elongation, both of which are components of the SAR. Separate signaling mechanisms were reported to operate downstream of phytochromes to regulate elongation and flowering responses to low R/FR in *Arabidopsis* (Franklin, 2008).

In the current study, leaflet area increased overall under the low R/FR environment, although with some exceptions. Baldissera et al. (2014) reported that shaded alfalfa plants have larger leaves; however, those of wild-type *Arabidopsis* seedlings display a leaf size decrease in response to low R/FR (Devlin et al., 1999). Increased leaf area under a low PAR environment has been observed in various plant species (Kremer and Kropff, 1999; James and Bell, 2000; Liao et al., 2006), and the common assumption is that increased leaf area will help increase light interception. Moreover, plant leaves grown under high PAR have lower photosynthetic pigment content than leaves grown under low PAR (Mielke and Schaffer, 2010). Under light deficit conditions, plants set a series of compensatory mechanisms, such as incremental increase of the photosynthetic pigment content, change of the leaf angle toward the light source, or an increase in leaf area, to achieve higher light absorption efficiency. The latter might be the case in most *Lens* genotypes in a low R/FR environment.

Reproductive Growth in *Lens* Genotypes was Mainly Affected by Light Intensity with an Interaction of Light Quality

Light is the main source of energy for carbon assimilation and growth in plants, therefore yield and biomass reductions occur under reduced light intensity (Baligar et al., 2006; Polthannee et al., 2011). In our study, the overall yield and above-ground biomass were highest under the high PAR environment, which indicates reproductive growth in *Lens* genotypes is mainly affected by light intensity. Low R/FR induced SARs involve a marked redirection of assimilates toward elongation and away from structures dedicated to resource acquisition and storage in natural conditions that limit water and nutrient resources (Smith and Whitelam, 1997). For crop species, a SAR could lead to

decreased yield if plants expend resources on vegetative growth at the expense of reproductive development (Kebrom and Brutnell, 2007; Casal, 2013). In the low R/FR indoor settings of our study, the SARs were clear in most *Lens* genotypes, resulting in reduced yield even though plants had sufficient water and nutrients. The *L. lamottei* group and *L. ervoides* IG 72646, however, maintained comparable yield, biomass, and harvest index under all three light environments, which may indicate better adaptation to changes in light environment. Maloof et al. (2001) reported natural variation in light sensitivity across a diverse set of *A. thaliana* accessions, and Hancock et al. (2011) detected and identified PAR-adaptive alleles in *A. thaliana* using a genome-wide scan. Identification of these light-adaptive alleles would further help on understanding of the genetic basis of light responses in lentil; such work is currently under way in our group.

CONCLUSION

Differences in light quality and intensity will affect the growth and development patterns of *Lens* plants, although some species are less affected than others. The high R/FR ratio created by fluorescent bulbs is not uncommon in controlled environment growth chambers. The results suggest that caution should be exercised in controlled environment growth chambers because the spectral property of the artificial light sources could severely delay the flowering of some crops, such as lentil, and thereby cause mismatch and delay for indoor breeding cycles. The identification of some *Lens* species that were less sensitive to R/FR related light quality and PAR related light intensity change may indicate a better adaptability toward changes in light environment. These varied responses might represent a source of genetic diversity that could be deployed in cultivated lentil to allow it to better handle sub-optimal light environments. Overall, increased understanding of light responses will help improve our ability to develop cultivars that have better adaptation to variable light environments.

AUTHOR CONTRIBUTIONS

Conception and design of the experiments: AV, SS, HY, KB. Acquisition of the data: SS, HY. Analysis and interpretation of the data: HY, KB, AV, SS. Drafting and revision of the paper: HY, KB, AV, SS.

FUNDING

This work was funded by Saskatchewan Pulse Growers (#BRE1202) and the NSERC Industrial Research Chair program (#IRCPG 395994-09).

ACKNOWLEDGMENTS

We acknowledge Mr. Adam Harrison of the Phytotron facility at the University of Saskatchewan for accommodating the required

light systems and Ms. Lacey-Anne Sanderson for helping with the heatmap presentations. Appreciation also goes to Ms. Devini De Silva, Ms. Chandra Bandara, and Ms. Brianna Jansen for helping with the maintenance of plants and experiments.

REFERENCES

- Baldissera, T. C., Frak, E., de Faccio Carvalho, P. C., and Louarn, G. (2014). Plant development controls leaf area expansion in alfalfa plants competing for light. *Ann. Bot.* 113, 145–157. doi: 10.1093/aob/mct251
- Baligar, V. C., Fageria, N. K., Paiva, A. Q., Silveira, A., de Souza, J. O. Jr., and Lucena, E. (2008). “Light intensity effects on growth and nutrient-use efficiency of tropical legume cover crops,” in *Toward Agroforestry Design: An Ecological Approach*, eds S. Jose and A. M. Gordon (Berlin: Springer), 67–79.
- Baligar, V. C., Fageria, N. K., Paiva, A. Q., Silveira, A., Pomella, A. W. V., and Machado, R. C. R. (2006). Light intensity effects on growth and micronutrient uptake by tropical legume cover crops. *J. Plant Nutr.* 29, 1959–1974. doi: 10.1080/01904160600927633
- Ballare, C. L., Scopel, A. L., and Sanchez, R. A. (1990). Far-red radiation reflected from adjacent leaves: an early signal of competition in plant canopies. *Science* 247, 329–332. doi: 10.1126/science.247.4940.329
- Ballare, C. L., Scopel, A. L., and Sanchez, R. A. (1991). Photocontrol of stem elongation in plant neighbourhoods: effects of photon fluence rate under natural conditions of radiation. *Plant Cell Environ.* 14, 57–65. doi: 10.1111/j.1365-3040.1991.tb01371.x
- Bethlenfalvay, G. B., and Phillips, D. (1977). Effect of light intensity on efficiency of carbon dioxide and nitrogen reduction in *Pisum sativum* L. *Plant Phys.* 60, 868–871. doi: 10.1104/pp.60.6.868
- Board, J. (2000). Light interception efficiency and light quality affect yield compensation of soybean at low plant populations. *Crop Sci.* 40, 1285–1294. doi: 10.2135/cropsci2000.4051285x
- Casal, J. J. (2013). Canopy light signals and crop yield in sickness and in health. *ISRN Agronomy* 2013:650439. doi: 10.1155/2013/650439
- Chen, M., Chory, J., and Fankhauser, C. (2004). Light signal transduction in higher plants. *Annu. Rev. Genet.* 38, 87–117. doi: 10.1146/annurev.genet.38.072902.092259
- Clack, T., Mathews, S., and Sharrock, R. A. (1994). The phytochrome apoprotein family in *Arabidopsis* is encoded by five genes: the sequences and expression of PHYD and PHYE. *Plant Mol. Biol.* 25, 413–427. doi: 10.1007/BF00043870
- Cubero, J. I., Perez de la Vega, M., and Frattini, R. (2009). “Origin, phylogeny, domestication and spread,” in *The Lentil: Botany, Production and Uses*, eds W. Erskine, F. Muehlbauer, A. Sarker, and B. Sharma (Wallingford, UK: CAB International), 13–33. doi: 10.1079/9781845934873.0013
- Devlin, P. F., Halliday, K. J., Harberd, N. P., and Whitelam, G. C. (1996). The rosette habit of *Arabidopsis thaliana* is dependent upon phytochrome action: novel phytochromes control internode elongation and flowering time. *Plant J.* 10, 1127–1134. doi: 10.1046/j.1365-313X.1996.10061127.x
- Devlin, P. F., Robson, P. R. H., Patel, S. R., Goosey, L., Sharrock, R. A., and Whitelam, G. C. (1999). Phytochrome D acts in the shade-avoidance syndrome in *Arabidopsis* by controlling elongation growth and flowering time. *Plant Physiol.* 119, 909–915. doi: 10.1104/pp.119.3.909
- Erskine, W., Muehlbauer, F. J., and Short, R. W. (1990). Stages of development in lentil. *Exp. Agric.* 26, 297–302. doi: 10.1017/S0014479700018457
- Evers, J. B., Van Der Krol, A. R., Vos, J., and Struik, P. C. (2011). Understanding shoot branching by modelling form and function. *Trends Plant Sci.* 16, 464–467. doi: 10.1016/j.tplants.2011.05.004
- Fiala, J. V., Tullu, A., Banniza, S., Séguin-Swartz, G., and Vandenberg, A. (2009). Interspecies transfer of resistance to anthracnose in lentil (*Lens culinaris* Medic.). *Crop Sci.* 49, 825–830. doi: 10.2135/cropsci2008.05.0260
- Franklin, K. A. (2008). Shade avoidance. *New Phytol.* 179, 930–944. doi: 10.1111/j.1469-8137.2008.02507.x
- Franklin, K. A., Praekelt, U., Stoddart, W. M., Billingham, O. E., Halliday, K. J., and Whitelam, G. C. (2003). Phytochromes B, D, and E act redundantly to control multiple physiological responses in *Arabidopsis*. *Plant Physiol.* 131, 1340–1346. doi: 10.1104/pp.102.015487
- Franklin, K. A., and Quail, P. H. (2010). Phytochrome functions in *Arabidopsis* development. *J. Exp. Bot.* 61, 11–24. doi: 10.1093/jxb/erp304
- Franklin, K. A., and Whitelam, G. C. (2005). Phytochromes and shade-avoidance responses in plants. *Ann. Bot.* 96, 169–175. doi: 10.1093/aob/mci165
- Frattini, R., Duran, Y., Garcia, P., and Perez de la Vega, M. (2007). Identification of quantitative trait loci (QTL) for plant structure, growth habit and yield in lentil. *Span. J. Agric. Res.* 5, 348–356. doi: 10.5424/sjar/2007053-255
- Gong, W. Z., Jiang, C. D., Wu, Y. S., Chen, H. H., Liu, W. Y., and Yang, W. Y. (2015). Tolerance vs. avoidance: two strategies of soybean seedlings in response to shade in intercropping. *Photosynthetica* 53, 259–268. doi: 10.1007/s11099-015-0103-8
- Halliday, K. J., Koornneef, M., and Whitelam, G. C. (1994). Phytochrome B and at least one other phytochrome mediate the accelerated flowering response of *Arabidopsis thaliana* L. to low red/far-red ratio. *Plant Physiol.* 104, 1311–1315. doi: 10.1104/pp.104.4.1311
- Han, Y. J., Song, P. S., and Kim, J. I. (2007). Phytochrome-mediated phytomorphogenesis in plants. *J. Plant Biol.* 50, 230–240. doi: 10.1007/BF03030650
- Hancock, A. M., Brachi, B., Faure, N., Horton, M. W., Jarymowycz, L. B., Sperone, F. G., et al. (2011). Adaptation to climate across the *Arabidopsis thaliana* genome. *Science* 334, 83–86. doi: 10.1126/science.1209244
- Hecht, V., Foucher, F., Ferrandiz, C., Macknight, R., Navarro, C., Morin, J., et al. (2005). Conservation of *Arabidopsis* flowering genes in model legumes. *Plant Physiol.* 137, 1420–1434. doi: 10.1104/pp.104.057018
- Heraut-Bron, V., Robin, C., Varlet-Grancher, C., Afif, D., and Guckert, A. (2000). Light quality (red:far-red ratio): does it affect photosynthetic activity, net CO₂ assimilation and morphology of young white clover leaves? *Can. J. Bot.* 77, 1425–1431. doi: 10.1139/cjb-77-10-1425
- James, S. A., and Bell, D. T. (2000). Influence of light availability on leaf structure and growth of two *Eucalyptus globulus* ssp. *Globulus* provenances. *Tree Physiol.* 20, 1007–1008. doi: 10.1093/treephys/20.15.1007
- Kebrom, T. H., and Brutnell, T. P. (2007). The molecular analysis of the shade avoidance syndrome in the grasses has begun. *J. Exp. Bot.* 58, 3079–3089. doi: 10.1093/jxb/erm205
- Kremer, E., and Kropff, M. J. (1999). Comparative growth of triazine-susceptible and -resistant biotypes of *Solanum nigrum* at different light levels. *Ann. Bot.* 83, 637–644. doi: 10.1007/BF00049537
- Li, J. G., Li, G., Wang, H. Y., and Deng, X. W. (2011). Phytochrome signaling mechanisms. *Arabidopsis Book* 9:e0148. doi: 10.1199/tab.0148
- Liao, J. X., Zou, X. Y., Ge, Y., and Chang, J. (2006). Effects of light intensity on growth of four *Mosla* species. *Bot. Stud.* 47, 403–408.
- Lorrain, S., Genoud, T., and Fankhauser, C. (2006). Let there be light in the nucleus. *Curr. Opin. Plant Biol.* 9, 509–514. doi: 10.1016/j.pbi.2006.07.013
- Maloof, J. N., Borevitz, J. O., Dabi, T., Lutes, J., Nehring, R. B., Redfern, J. L., et al. (2001). Natural variation in light sensitivity of *Arabidopsis*. *Nat. Genet.* 29, 441–446. doi: 10.1038/ng777
- Mielke, M. S., and Schaffer, B. (2010). Photosynthetic and growth responses of *Eugenia uniflora* L. seedlings to soil flooding and light intensity. *Environ. Exp. Bot.* 68, 113–121. doi: 10.1016/j.envexpbot.2009.11.007
- Mobini, S. H., Lulsdorf, M., Warkentin, T. D., and Vandenberg, A. (2016). Low red: far-red light ratio causes faster in vitro flowering in lentil. *Can. J. Plant Sci.* 96, 908–918. doi: 10.1139/cjps-2015-0282
- Mobini, S. H., Yuan, H. Y., Lulsdorf, M., and Vandenberg, A. (2012). “Photosynthesis and flowering enhancement affected by the action spectrum,” in *Proceedings of the 8th Canadian Plant Biotechnology Conference* (Guelph, ON: The University of Guelph).
- Möglich, A., Yang, X., Ayers, R. A., and Moffat, K. (2010). Structure and function of plant photoreceptors. *Annu. Rev. Plant Biol.* 61, 21–47. doi: 10.1146/annurev-arplant-042809-112259

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00386/full#supplementary-material>

- Morelli, G., and Ruberti, I. (2000). Shade avoidance responses: driving auxin along lateral routes. *Plant Physiol.* 122, 621–626. doi: 10.1104/pp.122.3.621
- Podder, R., Banniza, S., and Vandenberg, A. (2012). Screening of wild and cultivated lentil germplasm for resistance to stemphylium blight. *Plant Gen. Res.* 11, 26–35. doi: 10.1017/S1479262112000329
- Polthane, A., Promsaena, K., and Laoken, A. (2011). Influence of low light intensity on growth and yield of four soybean cultivars during wet and dry seasons of northeast Thailand. *Agric. Sci.* 2, 61–67. doi: 10.4236/as.2011.22010
- Quail, P. H. (2002). Phytochrome photosensory signaling networks. *Nat. Rev. Mol. Cell Biol.* 3, 85–93. doi: 10.1038/nrm728
- Rockwell, N. C., Su, Y. S., and Lagarias, J. C. (2006). Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* 57, 837–858. doi: 10.1146/annurev.arplant.56.032604.144208
- Runkle, E. S., and Heins, R. D. (2001). Specific functions of red, far-red and blue light in flowering and stem extension of long-day plants. *J. Am. Soc. Hortic. Sci.* 126, 275–282.
- Schmitt, J., Stinchcombe, J. R., Heschel, M. S., and Huber, H. (2003). The adaptive evolution of plasticity: phytochrome-mediated shade avoidance responses. *Integr. Comp. Biol.* 43, 459–469. doi: 10.1093/icb/43.3.459
- Sharrock, R. A., and Quail, P. H. (1989). Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. *Genes Dev.* 3, 1745–1757. doi: 10.1101/gad.3.11.1745
- Smith, H. (1982). Light quality, photoperception, and plant strategy. *Ann. Rev. Plant Physiol.* 33, 481–518. doi: 10.1146/annurev.pp.33.060182.002405
- Smith, H. (1994). “Sensing the light environment: the functions of the phytochrome family,” in *Photomorphogenesis in Plants*, 2nd Edn, eds R. E. Kendrick and G. H. M. Kronenberg (Dordrecht: Kluwer Academic Publishers), 377–416.
- Smith, H., and Whitelam, G. C. (1990). Phytochrome, a family of photoreceptors with multiple physiological roles. *Plant Cell Environ.* 13, 695–707. doi: 10.1111/j.1365-3040.1990.tb01084.x
- Smith, H., and Whitelam, G. C. (1997). The shade avoidance syndrome: multiple responses mediated by multiple phytochromes. *Plant Cell Environ.* 20, 840–844. doi: 10.1046/j.1365-3040.1997.d01-104.x
- Summerfield, R. J., Roberts, E. H., Erskine, W., and Ellis, R. H. (1985). Effects of temperature and photoperiod on flowering in lentils (*Lens culinaris* Medic). *Ann. Bot.* 56, 659–671. doi: 10.1093/oxfordjournals.aob.a087055
- Tullu, A., Banniza, S., Tar'an, B., Warkentin, T., and Vandenberg, A. (2010). Sources of resistance to ascochyta blight in wild species of lentil (*Lens culinaris* Medik.). *Genet. Resour. Crop Evol.* 57, 1053–1063. doi: 10.1007/s10722-010-9547-7
- Tullu, A., Buchwaldt, L., Lulsdorf, M., Banniza, S., Barlow, B., Slinkard, A. E., et al. (2006). Sources of resistance to anthracnose (*Colletotrichum truncatum*) in wild *Lens* species. *Genet. Resour. Crop Evol.* 53, 111–119. doi: 10.1007/s10722-004-1586-5
- Ueoka-Nakanishi, H., Hori, N., Ishida, K., Ono, N., Yamashino, T., Nakamishi, N., et al. (2011). Characterization of shade avoidance responses in *Lotus japonicas*. *Biosci. Biotechnol. Biochem.* 75, 2148–2154. doi: 10.1271/bbb.110442
- Vail, S., Streliaoff, J. V., Tullu, A., and Vandenberg, A. (2012). Field evaluation of resistance to *Colletotrichum truncatum* in *Lens culinaris*, *Lens ervoides*, and *Lens ervoides* x *Lens culinaris* derivatives. *Field Crops Res.* 126, 145–151. doi: 10.1016/j.fcr.2011.10.002
- Whitelam, G. C., and Devlin, P. F. (1997). Roles of different phytochromes in *Arabidopsis* photomorphogenesis. *Plant Cell Environ.* 20, 752–758. doi: 10.1046/j.1365-3040.1997.d01-100.x
- Whitelam, G. C., and Smith, H. (1991). Retention of phytochrome-mediated shade avoidance responses in phytochrome-deficient mutants of *Arabidopsis*, cucumber and tomato. *J. Plant Physiol.* 39, 119–125. doi: 10.1016/S0176-1617(11)80176-8
- Wong, M. M. L., Gujaria-Verma, N., Ramsay, L., Yuan, H. Y., Caron, C., Diapari, M., et al. (2015). Classification and characterization of species within the genus *Lens* using genotyping-by-sequencing (GBS). *PLoS ONE* 10:e0122025. doi: 10.1371/journal.pone.0122025
- Wu, F. Q., Zhang, X. M., Li, D. M., and Fu, Y. F. (2011). Ectopic expression reveals a conserved PHYB homolog in soybean. *PLoS ONE* 6:e27737. doi: 10.1371/journal.pone.0027737
- Yuan, H. Y., Saha, S., De Silva, D., Vandenberg, A., and Bett, K. E. (2015). “Flowering and growth responses of *Lens* genotypes to changes in light environment,” in *Proceedings of the North American Pulse Improvement Association Biennial Meeting (NAPIA) 2015*, Niagara Falls, ON.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Yuan, Saha, Vandenberg and Bett. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comprehensive Analysis of the Soybean (*Glycine max*) *GmLAX* Auxin Transporter Gene Family

Chenglin Chai[†], Yongqin Wang[†], Babu Valliyodan and Henry T. Nguyen^{*}

Division of Plant Sciences, National Center for Soybean Biotechnology, University of Missouri, Columbia, MO, USA

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica - Universidade Nova de
Lisboa, Portugal

Reviewed by:

Jeanne Marie Harris,
University of Vermont, USA
Cristina Ferrandiz,
Consejo Superior de Investigaciones
Científicas- Instituto de Biología
Molecular y Celular de Plantas, Spain

*Correspondence:

Henry T. Nguyen
nguyenhenry@missouri.edu

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 12 November 2015

Accepted: 22 February 2016

Published: 09 March 2016

Citation:

Chai C, Wang Y, Valliyodan B and
Nguyen HT (2016) Comprehensive
Analysis of the Soybean (*Glycine max*)
GmLAX Auxin Transporter Gene
Family. *Front. Plant Sci.* 7:282.
doi: 10.3389/fpls.2016.00282

The phytohormone auxin plays a critical role in regulation of plant growth and development as well as plant responses to abiotic stresses. This is mainly achieved through its uneven distribution in plant via a polar auxin transport process. Auxin transporters are major players in polar auxin transport. The *AUXIN RESISTENT 1/LIKE AUX1* (*AUX/LAX*) auxin influx carriers belong to the amino acid permease family of proton-driven transporters and function in the uptake of indole-3-acetic acid (IAA). In this study, genome-wide comprehensive analysis of the soybean *AUX/LAX* (*GmLAX*) gene family, including phylogenetic relationships, chromosome localization, and gene structure, was carried out. A total of 15 *GmLAX* genes, including seven duplicated gene pairs, were identified in the soybean genome. They were distributed on 10 chromosomes. Despite their higher percentage identities at the protein level, *GmLAX*s exhibited versatile tissue-specific expression patterns, indicating coordinated functioning during plant growth and development. Most *GmLAX*s were responsive to drought and dehydration stresses and auxin and abscisic acid (ABA) stimuli, in a tissue- and/or time point- sensitive mode. Several *GmLAX* members were involved in responding to salt stress. Sequence analysis revealed that promoters of *GmLAX*s contained different combinations of stress-related cis-regulatory elements. These studies suggest that the soybean *GmLAX*s were under control of a very complex regulatory network, responding to various internal and external signals. This study helps to identify candidate *GmLAX*s for further analysis of their roles in soybean development and adaption to adverse environments.

Keywords: soybean, auxin transporter, *GmLAX*, abiotic stress, drought, salinity, dehydration, abscisic acid

INTRODUCTION

The first discovered plant hormone, auxin (Went, 1926), is a key regulator of many aspects of plant growth and development, including embryogenesis, organogenesis, vascular tissue formation, and root and shoot tropisms (Petrasek and Friml, 2009; Swarup and Péret, 2012). In addition, auxin plays an important role in temporal coordination of plants' responses to abiotic stresses (Ha et al., 2013; Min et al., 2014). Auxin is mostly synthesized in developing parts of plants such as the shoot apex and developing leaves and seeds (Ljun et al., 2002). From its places of synthesis, auxin is transported throughout whole plant body where various developmental or responsive events occur, such as lateral root formation, apical dominance, leaf and flower development, and tropic growth in response to light and gravity (Petrasek and Friml, 2009).

The global distribution of auxin over the plant body was achieved by two distinct transportation pathways: long-distance, fast, non-polar transport through phloem, and slow, cell-to-cell polar transport (Michniewicz et al., 2007). The polar transport of auxin from cell to cell is mediated through the orchestration of auxin influx and efflux carriers, including *AUXIN RESISTENT 1/LIKE AUX1* (*AUX/LAX*) influx carriers (Swarup et al., 2004, 2008), *PIN-FORMED* (*PIN*) efflux carriers (Petrasek et al., 2006), and *P-GLYCOPROTEIN* (*PGP*) proteins (Cho et al., 2007; Cho and Cho, 2013). *PIN* proteins are typically polar-localized on either the plasma membrane or endoplasmic reticulum (ER), which enable them to lead the directions of auxin flow. *AUX/LAX*s encode multimembrane-spanning transmembrane proteins, and function in auxin uptake and intercellular auxin flow. They share similarities with amino acid transporters and form a plant-specific subclass within the amino acid/auxin permease super family (Young et al., 1999; Péret et al., 2012).

In *Arabidopsis*, *AUX/LAX* influx carriers include four members, *AUX1* and *LAX1-3*. Despite their high-similarity in sequences and conservation in biochemical function, each member of the *AUX/LAX* family exhibits distinct spatiotemporal expression patterns and works either independently or coordinately in various developmental events (Péret et al., 2012; Swarup and Péret, 2012). *AUX1*, working together with the auxin efflux carrier *PIN2* and *AXR4* (required for the correct localization of *AUX1* protein), plays a key role in root gravitropism (Swarup et al., 2001, 2005; Dharmasiri et al., 2006; Péret et al., 2012). Interestingly, though expressed in neighboring non-hair cells (but not in root hair cells), *AUX1* can regulate root hair development and maintain root hair polarity by working together with *PIN2* (Grebe et al., 2002; Jones et al., 2009). *LAX2* is essential for vascular development in cotyledon (Péret et al., 2012), and all *AUX/LAX* influx carriers control vascular patterning and xylem differentiation in plant (Fàbregas et al., 2015). *LAX3* and *AUX1* coordinately regulates lateral root (LR) development, with the former in LR emergence step and the latter in LR initiation step, respectively (Marchant et al., 2002; Swarup et al., 2008). In the zone competent for LR formation, positive feedback regulation of *AUX1* and down-regulation of *PIN3* and *PIN7* enhances the local auxin maxima, which leads to LR initiation and regulates longitudinal spacing of LRs (Laskowski et al., 2008). However, it was found that, though controlling normal LR frequency, *AUX1* and *PIN* transporters were not involved in mechanical curvature-elicited LR formation, where a Ca^{2+} -dependent signaling pathway was suggested to operate in parallel with and possibly interact with the auxin-dependent pathway (Richter et al., 2009). Evidence showed that phyllotactic patterning occurred through the teamwork of all *AUX/LAX* genes (Stieger et al., 2002; Bainbridge et al., 2008; Swarup and Péret, 2012). In addition, *AUX/LAX* genes were also implicated in apical hook development (Vandenbussche et al., 2010) and embryonic root cell organization and plant embryogenesis (Ugartechea-Chirino et al., 2010; Robert et al., 2015).

A growing body of evidence has demonstrate that *AUX/LAX* auxin transporters play roles in plant adaptation to variable environmental conditions. *AUX/LAX* genes were involved in biotic interactions, both pathogenic and symbiotic, such as

nodule formation in *Casuarina glauca* (Péret et al., 2007) and cyst nematode infection in *Arabidopsis* (Lee et al., 2011). In cotton, transcript profiling analysis revealed that two *AUX/LAX* auxin influx genes were significantly induced in anther by high-temperature stress in a high-temperature tolerant line, but not in a high-temperature sensitive line (Min et al., 2014). *OsAUX1*, which controls lateral root initiation, primary root and root hair elongation in rice (Yu et al., 2015; Zhao et al., 2015), was responsive to Cd stress (Yu et al., 2015), as well as to alkaline stress-mediated inhibition of root elongation (Li et al., 2015b). In *Arabidopsis*, *AUX1* and *PIN2* can protect LR formation under iron stress (Li et al., 2015a). Some members of the *AUX/LAX* gene family in sorghum and maize were in response to hormonal and abiotic stress treatments at transcriptional level (Shen et al., 2010; Yue et al., 2015).

Despite the remarkable progress in the model plant *Arabidopsis*, little is known about the auxin influx carriers in soybean. Soybean is one of the most economically important crops, being a major source of plant protein and oil as well as other beneficial chemicals for human (Chai et al., 2015). Understanding the role of soybean auxin influx carriers in plant growth, development, and response to environmental cues will help to facilitate our crop breeding process in order to increase soybean yield. Therefore, presented here is comprehensive information about soybean auxin influx carriers pertaining to their identification, chromosomal distribution, gene structure, tissue expression pattern, transcriptional response to auxin and abiotic stress, and promoter cis-regulatory element analysis, which could be useful for further study.

MATERIALS AND METHODS

Identification of the *AUX/LAX* Auxin Influx Carriers in Soybean and other Legumes

Putative *AUX/LAX* auxin influx carriers in soybean, common bean (*Phaseolus vulgaris*), and *Medicago truncatula* were identified by BLAST searches against the corresponding reference genomes at Phytozome v10.3 (<http://phytozome.jgi.doe.gov/pz/portal.html>) using the full-length protein sequences of all four *Arabidopsis thaliana* *AUX/LAX*s (*AtAUX1* and *AtLAX1-3*) as queries. Following the same approach, putative *LjLAX* members were found from the *Lotus japonicus* genome assembly build 2.5 (<http://www.kazusa.or.jp/lotus/>).

Phylogenetic Analysis and Chromosomal Mapping of *GmLAX*s

Full-length protein sequences of *AUX/LAX*s from soybean, common bean, *Medicago truncatula*, *Lotus japonicus*, *Arabidopsis*, rice, maize, and sorghum were downloaded from Phytozome v10.3 website. Multiple-sequence alignments of the full-length protein sequences of *AUX/LAX*s were performed using Clustal Omega (McWilliam et al., 2013), and the alignment result of *AUX/LAX*s was provided in **Supplementary File 1**. The phylogenetic tree was then constructed by using the maximum likelihood method with a bootstrap analysis of 1000 replicates and the JTT with Freqs. (+F) Substitution Model using

MEGA 5.2 (Tamura et al., 2011). Identification numbers of all AUX/LAXs protein sequences used in the phylogenetic analysis were listed in **Supplementary Table S1**.

Chromosomal distribution of *GmLAX*s was drawn from top to bottom on soybean chromosomes according to the position of genes in genome annotation. The circular map showing syntenic blocks of soybean chromosomes was made using the online software SyMAP (Soderlund et al., 2011). Gene pairs with over 90% and highest nucleotide sequence identities were considered as duplicated genes, which were analyzed by the Lasergene v7.1 (DNASTAR, Madison, USA).

Gene Structure, Protein Profile, and Promoter Analysis

Gene structures of *GmLAX*s were constructed by comparing the coding sequences with their corresponding genomic sequences using Gene Structure Display Server (GSDS) software (Guo et al., 2007). Transmembrane domains of *GmLAX*s were analyzed and visualized using TMHMM2 (Krogh et al., 2001). Protein subcellular localization was predicted by WoLF PSORT (Horton et al., 2007). Other protein profiles of *GmLAX*s, such as protein length, molecular weight (MW) and isoelectric point (PI), were analyzed by Lasergene v7.1. Promoter sequences of 2000 base pairs upstream from the putative translation start site (ATG) of *GmLAX*s were downloaded from the Phytozome (v10.3) website. Stress-related cis-regulatory elements (Yamaguchi-Shinozaki and Shinozaki, 2005; Mochida et al., 2009; Naika et al., 2013) were analyzed following the same method (Chai et al., 2015).

Plant Growth, Treatment, and Tissue Collection

The soybean cultivar Williams 82 was used for all treatments. Plants were grown in 4-gallon pots containing a 3:1 mixture of turf and sand in growth chamber under the condition of 28/20°C day/night temperature, 14/10 h light/dark photoperiod, 800 $\mu\text{mol m}^{-2} \text{s}^{-1}$ light intensity and 60% humidity. Abiotic stress and hormone treatments, and tissue collection were carried out as previously described (Tran et al., 2009; Chai et al., 2015; Wang et al., 2015). For mild and moderate drought treatments, the leaf water potentials were -7 bar and -13 bar, respectively, and they each had their own well-watered controls. Likewise, the IAA (50 μM) and ABA (150 μM) treatments had their own mock controls at each time point. The salt (250 mM NaCl) and dehydration treatments at all-time points shared the same non-treatment controls. Samples were collected in biological triplicates, frozen immediately in liquid nitrogen, and kept at -80°C until use.

RNA Isolation, Primer Designing, and qRT-PCR

RNA isolation, primer designing, qRT-PCR reactions, and data analyses were performed as previously described (Chai et al., 2015). Primer specificity was confirmed by blasting each primer sequence against the soybean genome and by electrophoresis. Three biological and two technical replications were used in all qPCR experiments. The soybean ubiquitin gene

(Glyma.20G141600) was used as an internal standard for all qRT-PCR analysis. Quantitative PCR data were analyzed by using the comparative CT method (Schmittgen and Livak, 2008) in Microsoft Excel 2013, and statistical significance of fold change of gene expression (treatment/non-treatment control) was assessed by ANOVA and/or Student's *t*-test analysis. The primers used for qPCR analyses are provided in **Supplementary Table S2**.

RESULTS

Genome-Wide Identification of AUX/LAX Genes from Soybean and other Legumes

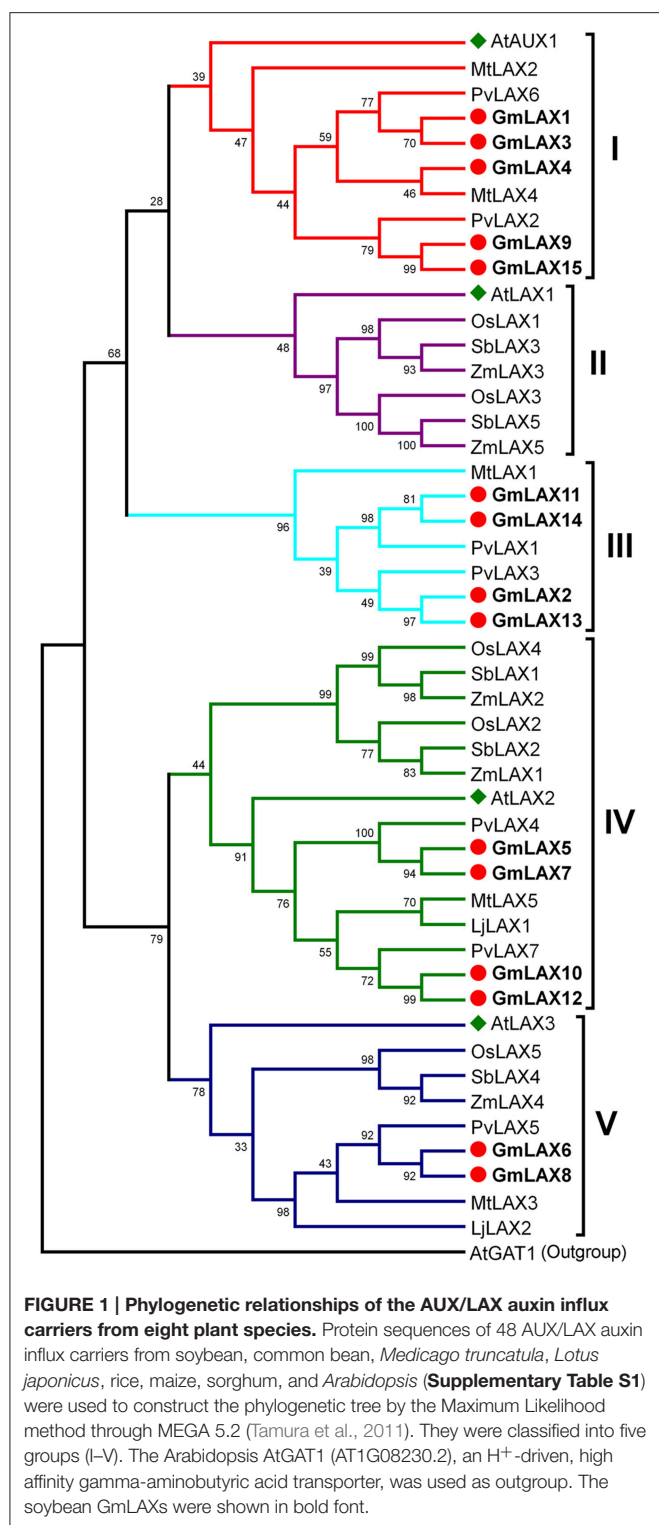
In order to explore the entire AUX/LAX gene family in soybean, BLAST searches against the soybean genome database (*Glycine max* Wm82.a2.v1) were conducted by using the *Arabidopsis* AUX/LAXs full-length protein sequences as queries. A total of 15 soybean *GmLAX*s were identified, which were designated as *GmLAX1* through *GmLAX15* according to their top-to-bottom positions on chromosomes from 1 to 18 (**Supplementary Table S3**), respectively. Using the same method, seven *PvLAX*s from common bean (*Phaseolus vulgaris* v1.0), five *MtLAX*s from *Medicago truncatula* (*Mt4.0v1*), and two *LjLAX*s from *Lotus japonicus* (genome assembly build 2.5) were identified. The soybean AUX/LAX gene family is expanded compared with other plant species. The number of AUX/LAXs is four in *Arabidopsis* (Péret et al., 2012), and five each in maize (Yue et al., 2015), rice (Shen et al., 2010; Zhao et al., 2015), and sorghum (Shen et al., 2010), respectively.

Phylogenetic Relationship of GmLAXs

Understanding the evolutionary relationships between *GmLAX*s and homologs from other plant species could be helpful in assessing their potential functions. Full-length protein sequences of 48 AUX/LAX genes from eight plant species, including four legumes (soybean, common bean, *Medicago truncatula*, and *Lotus japonicus*), three grasses (rice, maize and sorghum), and *Arabidopsis*, were used to construct the phylogenetic tree (**Figure 1**). The 48 AUX/LAX proteins were divided into five groups: I (AtAUX1-like, 10 members), II (AtLAX1-like, 7 members), III (legume-specific, 7 members), IV (AtLAX2-like, 15 members) and V (AtLAX3-like, 9 members). AUX/LAXs from the four legumes (or those from the three grasses) showed a very close phylogenetic relationship; while AUX/LAXs from the legumes and those from the grasses were evolved independently. The legume AUX/LAXs were classified into groups I, III, IV and V; while AUX/LAXs from the three grasses fell into groups II, IV, and V. There are five *GmLAX*s in the dicot-specific group I, four in the legume-specific group III, and four and two in group IV and V, respectively. The soybean *GmLAX*s showed the closest evolutionary relationships to the common bean *PvLAX*s. No AtLAX1 orthologs were found in soybean or other legumes.

Chromosomal Distribution, Gene Structure, and Protein Profiles

The 15 *GmLAX*s were unevenly distributed on 10 out of the 20 soybean chromosomes (**Figure 2**), with two *GmLAX*s each on chromosomes 3, 4, 6, 11, and 18, and one each on chromosomes



1, 2, 7, 12, and 14. The soybean genome has undergone two rounds of whole-genome duplication during its evolution (Schmutz et al., 2010), so it would be interesting to see whether gene duplication occurred in the *GmLAX* gene family. Analysis of nucleotide and amino acid identities of *GmLAX*s revealed seven

pair of duplicated genes, which shared over 95% identity at both the nucleotide and amino acid levels (Supplementary Table S4). Duplicated *GmLAX*s existed in the form of sister pairs in the phylogenetic tree (Figure 1), and they were linked together by lines in Figure 2A. The seven pairs of *GmLAX*s were all located in the duplication blocks on chromosomes (Figure 2B), indicating that they were formed during the most recent round of whole-genome duplication event.

All genes in the *GmLAX* family contained a conserved gene structure: eight exons and seven introns (Figure 3). *GmLAX* gene size varied greatly among members, mainly due to variations in intron sizes. Notably, duplicated genes or genes with closer evolutionary relationships had similar gene sizes. The encoded *GmLAX* proteins are of similar size, ranging from 465 to 506 amino acids. They shared other similar profiles, such as molecular weight and isoelectric point (Supplementary Table S3). Protein topology analysis revealed that all *GmLAX*s have a conserved core motif, which was composed of 10 transmembrane spanning domains (Supplementary Figure S1). Most of the *GmLAX*s were predicted to be plasma membrane-localized; while *GmLAX*2, 6, and 8 might be targeted to cytoplasm, and *GmLAX*13 targeted to both plasma membrane and cytoplasm (Supplementary Table S3).

Tissue/Organ-Specific Expression of *GmLAX*s

Gene functions are closely associated with where and how they are expressed. Transcript profiles of *GmLAX*s in seven tissues/organs (shoot apical meristem, flower, green pod, leaf, root, root tip, and nodule) were collected from soybean RNA-Seq data (Figure 4A, Libault et al., 2010). Gene expression patterns in root, stem, mature leaf, immature leaf, flower, pod, and seed at 14 and 21 days after flowering were studied using qRT-PCR (Figure 4B). Overall, the soybean *GmLAX*s showed very dynamic expression patterns. *GmLAX*5 and *GmLAX*7 were expressed at very low levels in almost all tissues. By contrast, *GmLAX*3 and *GmLAX*9 were expressed highly in most tissues. While for most *GmLAX*s, expressions were higher in some tissues/organs, but much lower or even barely detectable in others. For example, transcripts of *GmLAX*10 and *GmLAX*12 were higher in shoot apical meristem, root tip, and immature leaf, lower in root, stem, flower, and developing seeds, and almost undetectable in mature leaf, young pod and nodule. Another interesting scenario was that some duplicated genes, such as *GmLAX*10 and *GmLAX*12, and *GmLAX*9 and *GmLAX*15, exhibited similar expression patterns, but the expression levels in some tissues were quite different (Figure 4B).

Expression Profiles of *GmLAX*s under Abiotic Stresses

Soybean is one of the most drought and salinity sensitive crops. Its yield is significantly influenced by these abiotic stresses. In order to explore whether any *GmLAX* genes are involved in abiotic stress response, expressions of *GmLAX*s were investigated under drought, salinity, and dehydration using qRT-PCR (Figure 5A). Twelve *GmLAX*s were responsive to

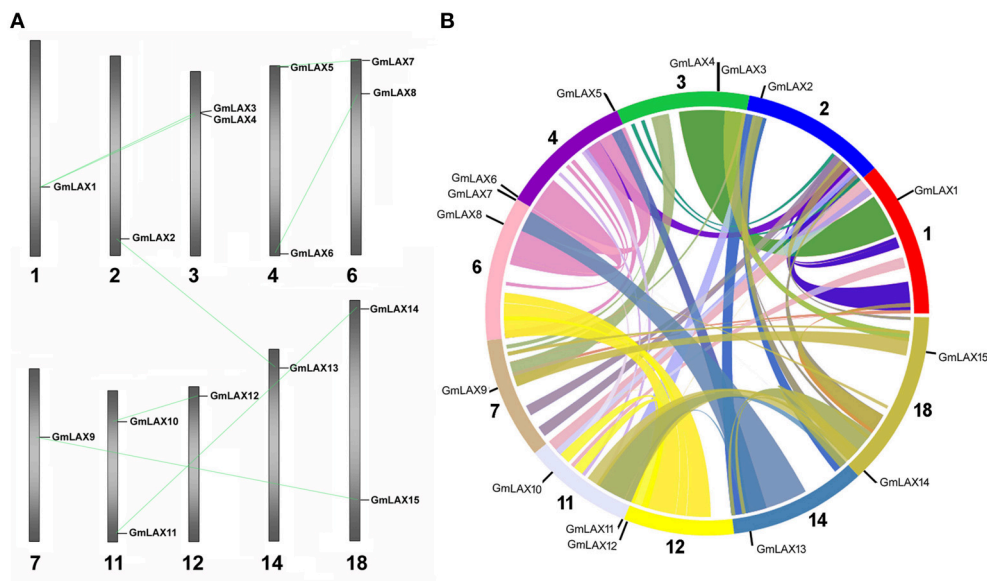


FIGURE 2 | Chromosomal distribution of the soybean *GmLAX*s. (A) Chromosomal locations of *GmLAX*s were shown from top to bottom on corresponding chromosomes (*Glycine max* Wm82.a2.v1). Duplicated genes are linked by gray lines. (B) A circular map of soybean chromosomes was drawn by SyMAP (<http://www.symapdb.org/>), showing the soybean *AUX/LAX*s localization in syntenic blocks.

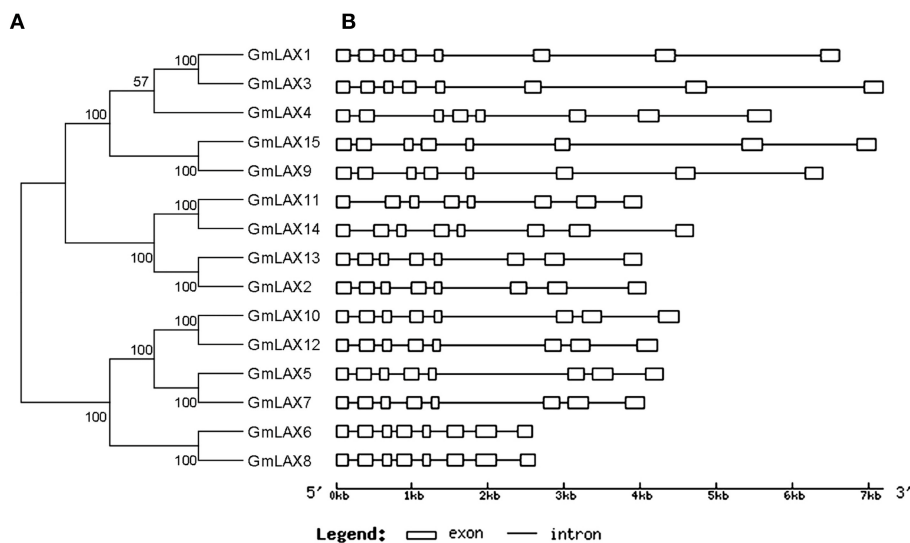


FIGURE 3 | Phylogenetic relationships (A) and gene structures (B) of *GmLAX*s. The phylogenetic analysis was carried out using MEGA 5.2 (Tamura et al., 2011). Gene exon-intron structures were made using the Gene Structure Display Server (Guo et al., 2007).

drought stresses. Under mild drought stress, six *GmLAX*s were transcriptionally regulated, with most of them (four out of six) being up-regulated in either shoot or root. While upon moderate drought stress, all 10 responsive *GmLAX*s were down-regulated, with two specifically in shoot, six solely in root, and two in both shoot and root. The response of *GmLAX*s to drought stresses was in a tissue-specific and stress magnitude-specific mode (Figure 5B). For example, seven *GmLAX*s were responsive to only one drought stress treatment (mild or moderate) and in

only one tissue (shoot or root). In some cases, the same gene was differentially regulated in different tissues. For instance, *GmLAX9* and *GmLAX15* were both up-regulated in shoots by mild drought stress but were down-regulated in roots by moderate drought stress (Figure 5A).

Quantitative-PCR analysis revealed that six *GmLAX*s were differentially expressed under salinity conditions and all of them were down-regulated (Figure 5). However, thirteen *GmLAX*s were down-regulated by dehydration. Of the 15 *GmLAX*s, three

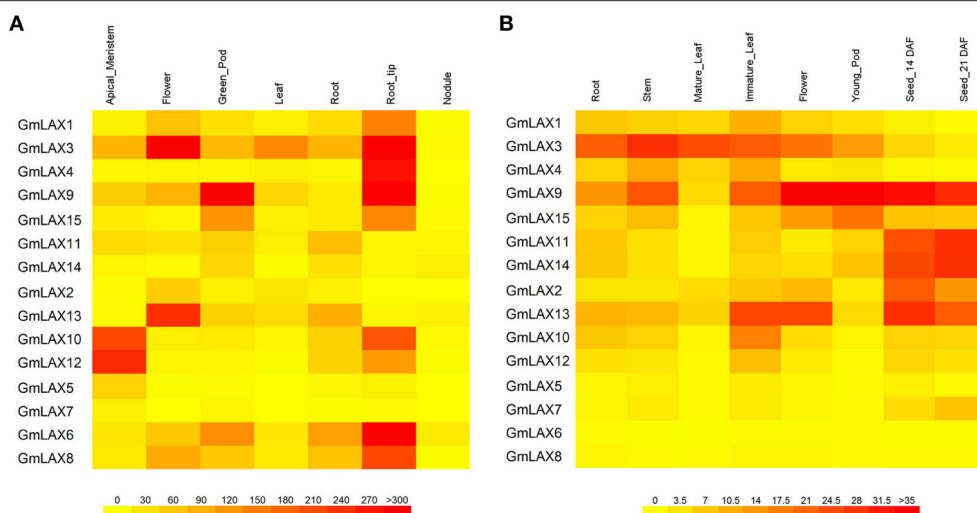


FIGURE 4 | Tissue/organ expression profiles of *GmLAX*s. (A) Expression of 15 *GmLAX*s in shoot apical meristem, flower, green pod, leaf, root, root tip, and nodule. RNA-Seq data (Libault et al., 2010) are shown as a heat map. (B) *GmLAX*s gene expression in root, stem, mature leaf, immature leaf, flower, pod, and seed at 14 and 21 days after flowering was analyzed by qRT-PCR. Relative expression values of *GmLAX*s were multiplied by 1000 and visualized as a heat map. All heat maps in this study were made by using the BAR Heatmapper (http://bar.utoronto.ca/ntools/cgi-bin/ntools_heatmapper.cgi).

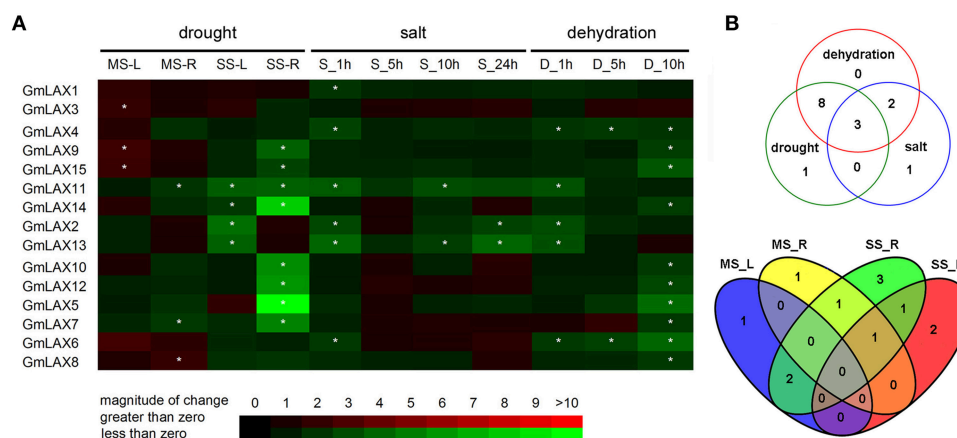


FIGURE 5 | Differential expression of *GmLAX*s in response to drought, salt, and dehydration. (A) Fold changes of *GmLAX*s gene expressions in shoots and roots under mild drought stress (–7 bar) and moderate drought stress (–13 bar) treatments, and in whole seedlings under salt (250 mM NaCl) and dehydration stress treatments. Data showed the means of three biological repeats in qRT-PCR analysis. Asterisks indicated fold changes (absolute value) ≥ 2 and $P < 0.05$ (Student's *t*-test). (B) Venn diagram analysis of data in (A). MS-L, mild drought stress-shoots; MS-R, mild drought stress-roots; SS-L, moderate drought stress-shoots; SS-R, moderate drought stress-roots; S-1 h, salt 1 h; S-5 h, salt 5 h; S-10 h, salt 10 h; S-24 h, salt 24 h; D-1 h, dehydration 1 h; D-5 h, dehydration 5 h; D-10 h, dehydration 10 h.

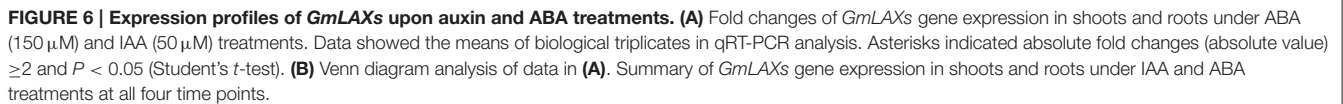
genes, i.e., *GmLAX11*, 2, and 13, were involved in responses to all three abiotic stresses, eight were exclusively responsive to drought and dehydration stresses, and two were specifically regulated by salt and dehydration stresses (Figure 5B). Only one gene was specifically responsive to salt stress, one to drought stress, and none of the 15 *GmLAX*s were dehydration-specific.

Expression Profiles of *GmLAX*s upon IAA and ABA treatment

Auxin is primarily regarded as a hormone that regulates plant growth and development, and also an effective regulator of auxin

carrier expression (Shen et al., 2010; Yue et al., 2015). As a stress hormone, ABA is involved in abiotic and biotic stress responses, and significant interactions between auxin and ABA signaling pathways have been well documented (Suzuki et al., 2001; Jain and Khurana, 2009; Anderson et al., 2012; Chen et al., 2014). In order to investigate whether the soybean *GmLAX*s were regulated by ABA and auxin, expression profiles of *GmLAX*s under treatments of these two hormones were analyzed by qRT-PCR (Figure 6A).

Twelve *GmLAX*s were differentially regulated by auxin at the transcriptional level, with four specifically in shoot, two in



Upon ABA treatment, 12 *GmLAXs* were differentially expressed, with two exclusively in shoot, four in root, and six in both tissues (**Figure 6**). Notably, most of the ABA-responsive *GmLAXs* were only responsive at certain time point(s) after ABA treatment, and most of the ABA-responsive genes in root were up-regulated. Expressions of *GmLAX3*, *10*, and *12* were induced by ABA in both shoots and roots (**Figure 6A**). Most interestingly, in both shoot and root, most auxin-responsive genes were also regulated by ABA (**Figure 6B**).

The versatile expression profiles of soybean *GmLAX* genes in different tissues/organs and in response to abiotic stresses and

Auxin is actively involved in various plant developmental processes (Petrasek and Friml, 2009; Swarup and Péret, 2012), as well as in plant responses to biotic and abiotic stresses (Kazan, 2013; Rahman, 2013). Control of these biological processes via auxin was achieved through its uneven distribution in plant, which was mainly mediated by coordinated actions of auxin influx and efflux transporters of three gene families: *AUX/LAX*, *PIN*, and *PGP* (Swarup and Péret, 2012). In this study, the

soybean *AUX/LAX* gene family was identified genome-wide and their expression profiles were analyzed.

GmLAXs Are Putative Auxin Influx Transporters in Soybean

In the present study, a total of 15 members (*GmLAX1* through *GmLAX15*) of the soybean *AUX/LAX* gene family were identified. The number of *LAX* genes in soybean is much larger than those from other plant species: three times the number of *LAX* genes in rice, maize, sorghum, or *Medicago truncatula*, and around twice the number in common bean (Figure 1). Although the absolute number of auxin influx transporter genes in soybean is much larger than those in other legume relatives, the ratio of *LAX* gene number to each genome size is comparable, indicating the expanded *GmLAX* gene family might be due to whole-genome duplication events during soybean evolution (Schmutz et al., 2010). This was further supported by gene duplication analysis and phylogenetic analysis (Figures 1, 2, and Supplementary Table S4), revealing seven pair of duplicated genes. The 48 *AUX/LAX*s from eight plant species were divided into five groups based on their phylogenetic relationships (Figure 1). The soybean *GmLAX*s in groups I, III, IV, and V might have experienced three, two, two and one round of duplication, respectively. However, the absence of legume *LAX*s from group II and the absence of grass *LAX*s from group I and III indicates gene loss from their ancestors during evolution.

Despite variance in gene and protein length, all *GmLAX*s exhibit a much conserved exon-intron organization with eight exons and seven introns (Figure 3). The gene structure of *LAX*s from other plants was less conserved compared to soybean (Stieger et al., 2002; Swarup et al., 2004, 2008; Kleine-Vehn et al., 2006; Bainbridge et al., 2008; Shen et al., 2010; Ugartechea-Chirino et al., 2010; Yue et al., 2015). Furthermore, all *GmLAX*s exhibited a conserved core motif with 10 transmembrane spanning domains (Supplementary Figure S1), suggesting that little has been changed in the protein structure of *GmLAX*s during evolution, probably due to its functional importance.

In spite of the significant conservation in gene and protein structure, expressions of *GmLAX*s at the transcription level among tissues/organs varied greatly (Figure 4). The high percent identity of duplicated genes at the protein level indicated that they might have conserved protein function as their *Arabidopsis* orthologs (Supplementary Table S1 and Figure 1). The tissue-specific expression profile analysis indicated that some duplicated gene pairs might play redundant roles in some tissues, such as *GmLAX10* and *GmLAX12* in shoot apical meristem, *GmLAX6* and *GmLAX8* in root tip, and *GmLAX11* and *GmLAX14* in developing seed, whereas only one copy of the duplicates might have function in some tissues, for instance *GmLAX3* in mature leaf, and *GmLAX9* in developing seed (Figures 1, 4). In *Arabidopsis*, all four *AUX/LAX* genes encode functional auxin influx carriers, but they have non-redundant and complementary expression profiles and play distinct functions: *AtAUX1* functioning in root gravitropism (Swarup et al., 2001), *LAX2* in vascular development and phyllotactic patterning by working together with *LAX1* (Péret et al., 2012; Swarup and Péret, 2012), and *LAX2* and *LAX3* coordinately regulating lateral

root development (Swarup et al., 2008). The soybean *LAX*s orthologs might play similar or very different roles during soybean development due to their versatile expression patterns. For example, the nine soybean *AUX1* (*Arabidopsis*) orthologs, which forms two sub-groups in the phylogenetic relationship analysis, with one sub-group containing *GmLAX1*, 3, 4, 9, and 15, and the other *GmLAX11*, 14, 2, and 13, exhibit very different tissue expression profiles between members within the same sub-group or from different sub-groups (Figures 1, 4). Further detailed cell-type specific expression pattern analysis of *GmLAX*s in different tissues/organs and during different developmental processes will help to determine their specific gene functions.

GmLAXs Were Responsive to Abiotic Stresses, and Auxin and ABA Hormonal Signals

Under abiotic stresses such as drought and salinity, plants usually first adaptively decrease growth rate before growth stops or death occurs. The uneven distribution of auxin, which is mainly mediated via auxin transporters, plays a key role in plants' adaptation to adverse conditions by adjusting growth rate. Crosstalk between auxin and biotic and abiotic stress signaling has been reported in some plant species (Ghanashyam and Jain, 2009; Jain and Khurana, 2009). In soybean, genome-wide transcriptome analyses showed that many hormone-related genes were differentially expressed in leaf and root under water deficit conditions (Le et al., 2012; Song et al., 2016). Most members of the soybean *PIN* gene family were responsive to various abiotic stresses and phytohormone stimuli (Wang et al., 2015). However, the precise molecular mechanism regarding regulation of auxin transport and distribution, which were achieved by coordination of different auxin transporters, is largely unknown in soybean. In this study, responses of *GmLAX*s, putative auxin influx carriers in soybean, to abiotic stresses and hormone signals, including auxin and ABA, were investigated (Figures 5, 6). Most *GmLAX*s were down-regulated by drought and dehydration, while only six *GmLAX*s were responsive to salt stress, and all of them were down-regulated. Decreased expression levels of *GmLAX*s might reflect down-regulation of auxin up-taking and/or transport, which might result in decreased or ceased growth of soybean sink tissues. This could at least partially explain the lower biomass and yield of soybean under abiotic stresses (Liu et al., 2003). The expression patterns of *GmLAX*s under abiotic stresses were different from those of the maize *AUX/LAX*s, which were up-regulated by salt and drought stresses in shoots, but were repressed in the roots (Yue et al., 2015). Interestingly, the sorghum *SbLAX*s exhibited irregular expression patterns in response to drought and salt (Shen et al., 2010). These studies suggested that the three plant species might have different mechanisms in responding to these unfavorable environments.

Auxin and ABA are two of the most important plant hormones, regulating plant growth and plant responses to environmental stresses, in both independent and coordinated manners. Most recently, several reports have indicated that auxin might mediate plant's adaptations to its adverse environment (Kazan, 2013; Rahman, 2013). Evidence suggests that auxin

transporters may play important roles during this process (Shen et al., 2010; Habets and Offringa, 2014; Yue et al., 2015). In *Arabidopsis*, ABA regulates root elongation through the activities of auxin and ethylene, likely operating in a linear pathway in this process (Thole et al., 2014), and ethylene inhibits root elongation through *AUX1* and auxin biosynthesis-related genes during alkaline stress (Li et al., 2015b). In soybean, auxin accumulation and distribution in the root altered upon abiotic stress and hormonal treatments, and some *GmPIN* genes likely contribute to auxin redistribution under these conditions (Wang et al., 2015). Therefore, auxin transporters might at least partially mediate the crosstalk between auxin, ABA and abiotic stresses. Our study revealed that many soybean *GmLAXs* were transcriptionally responsive to auxin and ABA stimuli (Figure 6). Interestingly, expressions of *GmLAX10* and *12* were induced by auxin and ABA in both root and shoot. The versatile expression responses of *GmLAXs* to the two hormones and abiotic stresses imply that these genes were under control of a very complex regulatory network. This was further supported by our analysis of stress-related cis-regulatory elements in promoters of *GmLAXs* (Supplementary Table S5). In response to various internal and external signals, the soybean *GmLAXs* might be actively involved in regulation of auxin distribution, thereby leading to plant growth adjustment and adaption to environmental stress conditions, by working together with other auxin transporters, such as *PINs* and *PGPs*.

Our study provides basic information on the soybean *GmLAX* gene family, and advances our knowledge on how these soybean auxin influx carriers function at the transcriptional level during plant development and adaption to adverse environments. This will help to identify candidates for further investigation and accelerate the research on abiotic stress tolerance mechanisms and development of soybean with improved plant performance.

REFERENCES

- Anderson, J. V., Dogramaci, M., Horvath, D. P., Foley, M. E., Chao, W. S., Suttle, J. C., et al. (2012). Auxin and ABA act as central regulators of developmental networks associated with paradormancy in Canada thistle (*Cirsium arvense*). *Funct. Integr. Genomics*. 12, 515–531. doi: 10.1007/s10142-012-0280-5
- Bainbridge, K., Guyomarc'h, S., Bayer, E., Swarup, R., Bennett, M., Mandel, T., et al. (2008). Auxin influx carriers stabilize phyllotactic patterning. *Genes. Dev.* 22, 810–823. doi: 10.1101/gad.462608
- Chai, C., Wang, Y., Joshi, T., Valliyodan, B., Prince, S., Michel, L., et al. (2015). Soybean transcription factor ORFeome associated with drought resistance: a valuable resource to accelerate research on abiotic stress resistance. *BMC Genomics* 16:596. doi: 10.1186/s12864-015-1743-6
- Chen, C., Twito, S., and Miller, G. (2014). New cross talk between ROS, ABA and auxin controlling seed maturation and germination unraveled in APX6 deficient *Arabidopsis* seeds. *Plant Signal. Behav.* 9:e976489. doi: 10.4161/15592324.2014.976489
- Cho, M., and Cho, H. (2013). The function of ABCB transporters in auxin transport. *Plant Signal. Behav.* 8:e22990. doi: 10.4161/psb.22990
- Cho, M., Lee, S. H., and Cho, H. T. (2007). P-glycoprotein4 displays auxin efflux transporter-like action in *Arabidopsis* root hair cells and tobacco cells. *Plant Cell*. 19, 3930–3943. doi: 10.1105/tpc.107.054288
- Dharmasiri, S., Swarup, R., Mockaitis, K., Dharmasiri, N., Singh, S. K., Kowalchuk, M., et al. (2006). AXR4 is required for localization of the auxin influx facilitator *AUX1*. *Science* 312, 1218–1220. doi: 10.1126/science.1122847
- Fàbregas, N., Formosa-Jordan, P., Confraria, A., Siligato, R., Alonso, J. M., Swarup, R., et al. (2015). Auxin influx carriers control vascular patterning and xylem differentiation in *Arabidopsis thaliana*. *PLoS Genet.* 11:e1005183. doi: 10.1371/journal.pgen.1005183
- Ghanashyam, C., and Jain, M. (2009). Role of auxin-responsive genes in biotic stress responses. *Plant Signal. Behav.* 4, 846–848. doi: 10.4161/psb.4.9.9376
- Grebe, M., Friml, J., Swarup, R., Ljung, K., Sandberg, G., Terlou, M., et al. (2002). Cell polarity signaling in *Arabidopsis* involves a BFA-sensitive auxin influx pathway. *Curr. Biol.* 12, 329–334. doi: 10.1016/S0960-9822(02)00654-1
- Guo, A. Y., Zhu, Q. H., Chen, X., and Luo, J. C. (2007). GSDS: a gene structure display server. *Yi chuan*. 29, 1023–1026. doi: 10.1360/yc-007-1023
- Ha, C. V., Le, D. T., Nishiyama, R., Watanabe, Y., Suliman, S., Tran, U. T., et al. (2013). The auxin response factor transcription factor family in soybean: genome-wide identification and expression analyses during development and water stress. *DNA Res.* 20, 511–524. doi: 10.1093/dnares/dst027
- Habets, M. E., and Offringa, R. (2014). PIN-driven polar auxin transport in plant developmental plasticity: a key target for environmental and endogenous signals. *New Phytol.* 203, 362–377. doi: 10.1111/nph.12831
- Horton, P., Park, K. J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., et al. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Res.* 35, W585–W587. doi: 10.1093/nar/gkm259

AUTHOR CONTRIBUTIONS

CC and YW designed and performed the experiments, analyzed data, and prepared the manuscript. BV and HN conceived and supervised the project and critically revised the manuscript. All authors have read, revised, and approved the manuscript.

ACKNOWLEDGMENTS

We thank Theresa A. Musket (University of Missouri) for her careful editing this manuscript. This research was funded by the Missouri Soybean Merchandising Council (Grant number 275F) and the United Soybean Board (Grant number 1204).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00282>

Supplementary File 1 | Clustal O (1.2.1) multiple sequence alignment of 48 AUX/LAX proteins from soybean, common bean, *Medicago truncatula*, *Lotus japonicus*, rice, maize, sorghum, and *Arabidopsis*.

Supplementary Figure S1 | Transmembrane helices of GmLAXs. Protein transmembrane topology was analyzed using the TMHMM Server v2.0 (Krogh et al., 2001).

Supplementary Table S1 | Protein sequences used in the phylogenetic relationship analysis.

Supplementary Table S2 | Primers used in qPCR analyses.

Supplementary Table S3 | GmLAXs gene information.

Supplementary Table S4 | Percent ORF nucleotide (bottom-left) and amino acid (up-right, bold) identities of GmLAXs.

Supplementary Table S5 | Analysis of stress-related cis-elements in the 2-kb promoter sequences of GmLAXs.

- Jain, M., and Khurana, J. P. (2009). Transcript profiling reveals diverse roles of auxin-responsive genes during reproductive development and abiotic stress in rice. *FEBS J.* 276, 3148–3162. doi: 10.1111/j.1742-4658.2009.07033.x
- Jones, A. R., Kramer, E. M., Knox, K., Swarup, R., Bennett, M. J., Lazarus, C. M., et al. (2009). Auxin transport through non-hair cells sustains root-hair development. *Nat. Cell Biol.* 11, 78–84. doi: 10.1038/ncb1815
- Kazan, K. (2013). Auxin and the integration of environmental signals into plant root development. *Ann. Bot.* 112, 1655–1665. doi: 10.1093/aob/mct229
- Kleine-Vehn, J., Dhonukshe, P., Swarup, R., Bennett, M., and Friml, J. (2006). Subcellular trafficking of the *Arabidopsis* auxin influx carrier AUX1 uses a novel pathway distinct from PIN1. *Plant Cell.* 18, 3171–1381. doi: 10.1105/tpc.106.042770
- Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Bio.* 305, 567–580. doi: 10.1006/jmbi.2000.4315
- Laskowski, M. I., Grieneisen, V. A., Hofhuis, H., Hove, C. A., Hogeweg, P., Marée, A. F., et al. (2008). Root system architecture from coupling cell shape to auxin transport. *PLoS Biol.* 6:e307. doi: 10.1371/journal.pbio.0060307
- Le, D. T., Nishiyama, R., Watanabe, Y., Tanaka, M., Seki, M., le Ham, H., et al. (2012). Differential gene expression in soybean leaf tissues at late developmental stages under drought stress revealed by genome-wide transcriptome analysis. *PLoS ONE* 7:e49522. doi: 10.1371/journal.pone.0049522
- Lee, C., Chronis, D., Kenning, C., Péret, B., Hewezi, T., Davis, E. L., et al. (2011). The novel cyst nematode effector protein 19C07 interacts with the *Arabidopsis* auxin influx transporter LAX3 to control feeding site development. *Plant Physiol.* 155, 866–880. doi: 10.1104/pp.110.167197
- Li, G., Song, H., Li, B., Kronzucker, H. J., and Shi, W. (2015a). Auxin resistant1 and PIN-FORMED2 protect sateral root formation in *Arabidopsis* under iron stress. *Plant Physiol.* 169, 2608–2623. doi: 10.1104/pp.15.00904
- Li, J., Xu, H. H., Liu, W. C., Zhang, X. W., and Lu, Y. T. (2015b). Ethylene inhibits root elongation during alkaline stress through AUXIN1 and associated changes in auxin accumulation. *Plant Physiol.* 168, 1777–1791. doi: 10.1104/pp.15.00523
- Libault, M., Farmer, A., Joshi, T., Takahashi, K., Langley, R. J., Franklin, L. D., et al. (2010). An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. *Plant J.* 63, 86–99. doi: 10.1111/j.1365-313x.2010.04222.x
- Liu, F., Andersen, M. N., and Jensen, C. R. (2003). Loss of pod set caused by drought stress is associated with water status and ABA content of reproductive structures in soybean. *Funct. Plant Biol.* 30, 271–280. doi: 10.1071/EP02185
- Ljun, K., Hul, A. K., Kowalczyk, M., Marchant, A., Celenza, J., Cohen, J. D., et al. (2002). Biosynthesis, conjugation, catabolism and homeostasis of indole-3-acetic acid in *Arabidopsis thaliana*. *Plant Mol. Biol.* 50, 309–332. doi: 10.1023/A:1016024017872
- Marchant, A., Bhalarao, R., Casimiro, I., Eklof, J., Casero, P. J., Bennett, M., et al. (2002). AUX1 promotes lateral root formation by facilitating indole-3-acetic acid distribution between sink and source tissues in the *Arabidopsis* seedling. *Plant Cell.* 14, 589–597. doi: 10.1105/tpc.010354
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., et al. (2013). Analysis tool web services from the EMBL-EBI. *Nucleic Acids Res.* 41, W597–W600. doi: 10.1093/nar/gkt376
- Michniewicz, M., Brewer, P. B., and Friml, J. I. (2007). Polar auxin transport and asymmetric auxin distribution. *Arabidopsis Book.* 5:e0108. doi: 10.1199/tab.0108
- Min, L., Li, Y., Hu, Q., Zhu, L., Gao, W., Wu, Y., et al. (2014). Sugar and auxin signaling pathways respond to high-temperature stress during anther development as revealed by transcript profiling analysis in cotton. *Plant Physiol.* 164, 1293–1308. doi: 10.1104/pp.113.232314
- Mochida, K., Yoshida, T., Sakurai, T., Yamaguchi-Shinozaki, K., Shinozaki, K., and Tran, L. S. (2009). *In silico* analysis of transcription factor repertoire and prediction of stress responsive transcription factors in soybean. *DNA Res.* 16, 353–369. doi: 10.1093/dnares/dsp023
- Naika, M., Shameer, K., Mathew, O. K., Gowda, R., and Sowdhamini, R. (2013). STIFDB2: an updated version of plant stress-responsive transcription factor database with additional stress signals, stress-responsive transcription factor binding sites and stress-responsive genes in *Arabidopsis* and rice. *Plant Cell Physiol.* 54:e8(1–15). doi: 10.1093/pcp/pcs185
- Péret, B., Swarup, K., Ferguson, A., Seth, M., Yang, Y., Dhondt, S., et al. (2012). AUX/LAX genes encode a family of auxin influx transporters that perform distinct functions during *Arabidopsis* development. *Plant Cell.* 24, 2874–2885. doi: 10.1105/tpc.112.097766
- Péret, B., Swarup, R., Jansen, L., Devos, G., Auguy, F., Collin, M., et al. (2007). Auxin influx activity is associated with *Frankia* infection during actinorhizal nodule formation in *Casuarina glauca*. *Plant Physiol.* 144, 1852–1862. doi: 10.1104/pp.107.101337
- Petrasek, J., and Friml, J. (2009). Auxin transport routes in plant development. *Development.* 136, 2675–2688. doi: 10.1242/dev.030353
- Petrasek, J., Mravec, J., Bouchard, R., Blakeslee, J. J., Abas, M., Seifertova, D., et al. (2006). PIN proteins perform a rate-limiting function in cellular auxin efflux. *Science* 312, 914–918. doi: 10.1126/science.1123542
- Rahman, A. (2013). Auxin: a regulator of cold stress response. *Physiol. Plant* 147, 28–35. doi: 10.1111/j.1399-3054.2012.01617.x
- Richter, G. L., Monshausen, G. B., Krol, A., and Gilroy, S. (2009). Mechanical stimuli modulate lateral root organogenesis. *Plant Physiol.* 151, 1855–1866. doi: 10.1104/pp.109.142448
- Robert, H. S., Grunewald, W., Sauer, M., Cannoot, B., Soriano, M., Swarup, R., et al. (2015). Plant embryogenesis requires AUX/LAX-mediated auxin influx. *Development* 142, 702–711. doi: 10.1242/dev.115832
- Schmittgen, T. D., and Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* 3, 1101–1108. doi: 10.1038/nprot.2008.73
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Shen, C., Bai, Y., Wang, S., Zhang, S., Wu, Y., Chen, M., et al. (2010). Expression profile of PIN, AUX/LAX and PGP auxin transporter gene families in *Sorghum bicolor* under phytohormone and abiotic stress. *FEBS J.* 277, 2954–2969. doi: 10.1111/j.1742-4658.2010.07706.x
- Soderlund, C., Bomhoff, M., and Nelson, W. M. (2011). SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res.* 39:e68. doi: 10.1093/nar/gkr123
- Song, L., Prince, S., Valliyodan, B., Joshi, T., Maldonado Dos Santos, J. V., Wang, J., et al. (2016). Genome-wide transcriptome analysis of soybean primary root under varying water-deficit conditions. *BMC Genomics* 17:57. doi: 10.1186/s12864-016-2378-y
- Stieger, P. A., Reinhardt, D., and Kuhlmeier, C. (2002). The auxin influx carrier is essential for correct leaf positioning. *Plant J.* 32, 509–517. doi: 10.1046/j.1365-313X.2002.01448.x
- Suzuki, M., Kao, C. Y., Cocciolone, S., and McCarty, D. R. (2001). Maize VP1 complements *Arabidopsis abi3* and confers a novel ABA/auxin interaction in roots. *Plant J.* 28, 409–418. doi: 10.1046/j.1365-313X.2001.01165.x
- Swarup, K., Benkova, E., Swarup, R., Casimiro, I., Péret, B., Yang, Y., et al. (2008). The auxin influx carrier LAX3 promotes lateral root emergence. *Nat. Cell Biol.* 10, 946–954. doi: 10.1038/ncb1754
- Swarup, R., Friml, J., Marchant, A., Ljung, K., Sandberg, G., Palme, K., et al. (2001). Localization of the auxin permease AUX1 suggests two functionally distinct hormone transport pathways operate in the *Arabidopsis* root apex. *Genes Dev.* 15, 2648–2653. doi: 10.1101/gad.210501
- Swarup, R., Kargul, J., Marchant, A., Zadik, D., Rahman, A., Mills, R., et al. (2004). Structure-function analysis of the presumptive *Arabidopsis* auxin permease AUX1. *Plant Cell.* 16, 3069–3083. doi: 10.1105/tpc.104.024737
- Swarup, R., Kramer, E. M., Perry, P., Knox, K., Leyser, H. M., Haseloff, J., et al. (2005). Root gravitropism requires lateral root cap and epidermal cells for transport and response to a mobile auxin signal. *Nat. Cell Biol.* 7, 1057–1065. doi: 10.1038/ncb1316
- Swarup, R., and Péret, B. (2012). AUX/LAX family of auxin influx carriers—an overview. *Front Plant Sci.* 3:225. doi: 10.3389/fpls.2012.00225
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Thole, J. M., Beisner, E. R., Liu, J., Venkova, S. V., and Strader, L. C. (2014). Absciscic acid regulates root elongation through the activities of auxin and ethylene in *Arabidopsis thaliana*. *G3 (Bethesda)* 4, 1259–1274. doi: 10.1534/g3.114.011080

- Tran, L. S., Quach, T. N., Guttikonda, S. K., Aldrich, D. L., Kumar, R., Neelakandan, A., et al. (2009). Molecular characterization of stress-inducible *GmNAC* genes in soybean. *Mol. Genet. Genomics*. 281, 647–664. doi: 10.1007/s00438-009-0436-8
- Ugartechea-Chirino, Y., Swarup, R., Swarup, K., Péret, B., Whitworth, M., Bennett, M., et al. (2010). The AUX1 LAX family of auxin influx carriers is required for the establishment of embryonic root cell organization in *Arabidopsis thaliana*. *Ann. Bot.* 105, 277–289. doi: 10.1093/aob/mcp287
- Vandenbussche, F., Petrasek, J., Zadnikova, P., Hoyerova, K., Pesek, B., Raz, V., et al. (2010). The auxin influx carriers AUX1 and LAX3 are involved in auxin-ethylene interactions during apical hook development in *Arabidopsis thaliana* seedlings. *Development*. 137, 597–606. doi: 10.1242/dev.040790
- Wang, Y., Chai, C., Valliyodan, B., Maupin, C., Annen, B., and Nguyen, H. T. (2015). Genome-wide analysis and expression profiling of the *PIN* auxin transporter gene family in soybean (*Glycine max*). *BMC Genomics* 16:951. doi: 10.1186/s12864-015-2149-1
- Went, F. (1926). On growth-accelerating substances in the coleoptile of *Avena sativa*. *Proc. K. Ned. Akad. Wet.* 30, 10–19.
- Yamaguchi-Shinozaki, K., and Shinozaki, K. (2005). Organization of cis-acting regulatory elements in osmotic- and cold-stress-responsive promoters. *Trends Plant Sci.* 10, 88–94. doi: 10.1016/j.tplants.2004.12.012
- Young, G. B., Jack, D. L., Smith, D. W., and Saier, M. H. Jr. (1999). The amino acid/auxin:proton symport permease family. *Biochim. Biophys. Acta* 1415, 306–322. doi: 10.1016/S0005-2736(98)00196-5
- Yu, C., Sun, C., Shen, C., Wang, S., Liu, F., Liu, Y., et al. (2015). The auxin transporter, OsAUX1, is involved in primary root and root hair elongation and in Cd stress responses in rice (*Oryza sativa* L.). *Plant J.* 83, 818–830. doi: 10.1111/tpj.12929
- Yue, R., Tie, S., Sun, T., Zhang, L., Yang, Y., Qi, J., et al. (2015). Genome-wide identification and expression profiling analysis of *ZmPIN*, *ZmPILS*, *ZmLAX* and *ZmABCB* auxin transporter gene families in maize (*Zea mays* L.) under various abiotic stresses. *PLoS ONE* 10:e0118751. doi: 10.1371/journal.pone.0118751
- Zhao, H., Ma, T., Wang, X., Deng, Y., Ma, H., Zhang, R., et al. (2015). OsAUX1 controls lateral root initiation in rice (*Oryza sativa* L.). *Plant Cell Environ.* 38, 2208–2222. doi: 10.1111/pce.12467

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Chai, Wang, Valliyodan and Nguyen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



GmAGL1, a MADS-Box Gene from Soybean, Is Involved in Floral Organ Identity and Fruit Dehiscence

Yingjun Chi^{1,2}, Tingting Wang¹, Guangli Xu¹, Hui Yang¹, Xuanrui Zeng¹, Yixin Shen², Deyue Yu¹ and Fang Huang^{1*}

¹ National Center for Soybean Improvement, Key Laboratory of Biology and Genetic Improvement of Soybean Ministry of Agriculture P.R. China, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China, ² College of Agro-grassland Science, Nanjing Agricultural University, Nanjing, China

OPEN ACCESS

Edited by:

Susana Araújo,
Instituto de Tecnologia Química e
Biológica António Xavier
(ITQB-NOVA), Portugal

Reviewed by:

Ana Maria Rocha De Almeida,
University of California, Berkeley, USA
Jiayin Pang,
University of Western Australia,
Australia
Autar Krishen Matttoo,
United States Department
of Agriculture, USA

*Correspondence:

Fang Huang
fhuang@njau.edu.cn

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 01 July 2016

Accepted: 27 January 2017

Published: 09 February 2017

Citation:

Chi Y, Wang T, Xu G, Yang H,
Zeng X, Shen Y, Yu D and Huang F
(2017) GmAGL1, a MADS-Box Gene
from Soybean, Is Involved in Floral
Organ Identity and Fruit Dehiscence.
Front. Plant Sci. 8:175.
doi: 10.3389/fpls.2017.00175

MADS-domain proteins are important transcription factors involved in many aspects of plant reproductive development. In this study, a MADS-box gene, *Glycine max* AGAMOUS-LIKE1 (*GmAGL1*), was isolated from soybean flower. The transcript of *GmAGL1* was expressed in flowers and pods of different stages in soybean and was highly expressed in carpels. *GmAGL1* is a nucleus-localized transcription factor and can interact directly with SEP-like proteins in soybean flowers. Ectopic overexpression of *GmAGL1* resulted in the absence of petals in *Arabidopsis*. Moreover, morphological changes in the valves were observed in 35S:*GmAGL1 Arabidopsis* fruits that dehisced before the seeds reached full maturity. *GmAGL1* was found to be sufficient to activate the expression of *Arabidopsis* *ALC*, *IND*, *STK*, *SEP1*, and *SEP3*. Therefore, our data suggest that *GmAGL1* may play important roles in both floral organ identity and fruit dehiscence.

Keywords: AGAMOUS-LIKE1, soybean, fruit dehiscence, floral organ identity genes, protein interactions

INTRODUCTION

Angiosperms, or flowering plants, are the most diverse and numerous groups of plants. Despite their diversity, the most remarkable features distinguishing them from gymnosperms are their flowers and the production of fruits that contain seeds. Most flowers are composed of four types of floral organs with external sterile organs (sepals and petals) surrounding the reproductive structures (stamens and carpels) located in the center (Robles and Pelaz, 2005). The fertilized carpels give rise to the fruits, which protect the developing seeds and ultimately disperse the mature seeds into the environment. The vast majority of crops in the world are angiosperms, therefore, the regulation of flower and fruit development directly affects the economic benefits in agricultural production. Understanding the genetic factors regulating flower and fruit patterning may help to improve crop breeding.

Flower development has been the subject of intensive study for over the past 20 years. These studies led to the establishment of the well-known ABCDE model that explained the genetic regulation of floral organ identity determination. This model proposes that five classes of genes (A, B, C, D, and E) act in a combinatorial way to specify the distinct floral organs (Theissen and Saedler, 2001). The A+E protein complex determines sepals, the A+B+E complex specifies petals, the B+C+E complex specifies stamens, the C+E complex specifies carpels and the D+E

Abbreviations: BD, binding domain; GFP, green fluorescent protein; ORF, open reading frame; Y2H, yeast two hybrid.

complex determines the ovules (Colombo et al., 1995; Theissen and Saedler, 2001). In *Arabidopsis*, the class A genes are *APETALA1/2* (*API/2*) (Mandel et al., 1992; Jofuku et al., 1994); the class B genes are *AP3* (Jack et al., 1992) and *PISTILLATA* (*PI*) (Goto and Meyerowitz, 1994); the class C gene is *AGAMOUS* (*AG*) (Yanofsky et al., 1990); the class D genes are *SEEDSTICK* (*STK*) and *SHATTERPROOF1/2* (*SHP1/2*) (Favaro et al., 2003); and the class E genes are *SEPALLATA1/2/3/4* (*SEP1/2/3/4*) (Pelaz et al., 2000). Interestingly, all of these floral organ identity genes except *AP2* belong to the MADS-box gene family. Moreover, orthologs of these genes have been found in many other species, such as other eudicots, monocots, and even in gymnosperms (Robles and Pelaz, 2005; Arora et al., 2007; Katahata et al., 2014).

The genetic networks controlling fruit patterning are well-characterized in many plants (White, 2002; Xue et al., 2012; Karlova et al., 2014). In the model plant *Arabidopsis*, the fruit is a dry pod derived from two fused carpels called a silique. During dry fruit development, one main process is the differentiation of tissues required for dehiscence (Robles and Pelaz, 2005). Many genes related to fruit dehiscence have been identified. The MADS-box transcription factors *SHP1* and *SHP2* (previously named *AGL1* and *AGL5*) are the primary regulators of dry fruit dehiscence (Liljegren et al., 2000). Although *SHP* genes have been shown to specify carpel identity in a transcriptional complex (Favaro et al., 2003), they are best known for their functions in the differentiation of the dehiscence zone and the lignification of adjacent cells (Liljegren et al., 2000). Mutations in the *SHP* genes lead to indehiscent fruits, thus, inhibiting the seed dispersal process. *SHP* genes act at the top of the genetic regulatory hierarchy in valve margin formation and positively regulate *INDEHISCENT* (*IND*) and *ALCATRAZ* (*ALC*), which are two bHLH transcription factors also required for correct valve margin development (Rajani and Sundaresan, 2001; Liljegren et al., 2004). The expression of *SHP1*, *SHP2*, *IND*, and *ALC* in the valve margins is negatively controlled by *FRUITFUL* (*FUL*). As a MADS-box transcription factor, *FUL* is responsible for both ovary growth and valve cell differentiation (Ferrandiz et al., 2000; Liljegren et al., 2004).

Soybean (*Glycine max* [L.] Merr.) is an economically important global crop and it is now the main source of vegetable oil and protein. The development of flowers and pods directly affects seed yield and quality. Additionally, soybean is a self-pollinated crop, the anther and stigma of which are enclosed by the wing flap and keel. This floral morphology causes difficulties in cross-pollination and prevents crossing within or between individuals, which is unfavorable for soybean hybrid breeding (Huang et al., 2014). Therefore, genetic modification of the perianth is a priority in soybean hybrid breeding. The studies on MADS-box genes related to floral morphology will facilitate the production of valuable plant materials that have potential applications in soybean hybrid breeding. Based on a previous study on gene expression analysis by Affymetrix Gene Chip (Huang et al., 2009), a *Glycine max* *AGAMOUS-LIKE1* gene (Probe ID: Gma.11881.1.A1_at) was found to predominantly accumulate in soybean flowers and pods, indicating its role in floral organs and fruit development (Ma et al., 1991; Flanagan et al., 1996). We isolated *GmAGL1* from soybean flower and

characterized its spatial and temporal expression patterns. As a MADS domain protein, *GmAGL1* is a nucleus-localized transcription factor and functions in a multimeric complex with SEP-like proteins. *GmAGL1* was sufficient to activate the expression of *ALC*, *IND*, *STK*, *SEP1*, and *SEP3* in *Arabidopsis*. The ectopic overexpression of *GmAGL1* resulted in an abnormal floral organ identity in *Arabidopsis*, in which the petals were completely absent. Moreover, morphological changes in the valves were observed in 35S:*GmAGL1* fruits that dehisced before the seeds reached full maturity.

MATERIALS AND METHODS

Plant Materials and Growth

Soybean plants (cv. Jackson) were grown under field conditions in Nanjing Agricultural University, China. Young leaves, roots, stems and shoot apical meristems (SAMs) were collected at the third euphylls expanding stage. Flowers were harvested at different stages from tiny buds to mature flowers. Four types of floral organs were obtained from mature flowers. The developing pods were harvested at 5, 10, 20, 30, and 40 DAF (day after flowering). Seeds at 20 DAF were dissected to collect the seed coat, embryo and cotyledon under a dissection microscope with surgical blades and tweezers. All of these samples were lyophilized and stored at -80°C until used.

Arabidopsis thaliana, ecotype Columbia-0, was used for *GmAGL1* ectopic expression experiments. All plants were grown in a growth room at 22°C with 16 h light/8 h dark.

Isolation of *GmAGL1* Gene by RACE

Based on the sequence information from NCBI¹, the fragment (GenBank accession No: AW433203) was found to have an incomplete ORF. The rapid amplification of cDNA using the end (RACE) technique was employed to identify both the 3'- and 5'-ends of *GmAGL1* cDNA with a SMARTTM RACE cDNA Amplification Kit (Clontech, USA). The 3'-end was amplified from flower cDNA by two nested PCR reactions, using the gene specific primer 3'-GSP (5'-GAGAAAGCACAACAACGGCAACAG-3') for the first round PCR and another gene specific primer 3'-NGSP (5'-GCGAGTCAACCATACCTCCA-3') for nested PCR. For 5'-end amplification, the two primers used were 5'-GSP (5'-TGGAGGTATGGTTGACTCGCACA-3') and 5'-NGSP (5'-CCGTTGTTGTGCTTTCTCGTGTTC-3'). All the PCR products were gel purified, cloned into pGEM[®]-T easy vectors (Promega, USA) and sequenced (Invitrogen, Shanghai). According to the RACE results, the primer pair (sense: 5'-GCATAACACCAAAGAAGTAC-3' and anti-sense: 5'-TCACGAAACATAGGACGATT-3') was used to isolate the intact ORF of *GmAGL1* by RT-PCR.

Sequence and Phylogenetic Analysis

The ORF of *GmAGL1* and its deduced protein sequence were analyzed by BioXM (ver 2.6). Conserved domains were

¹<http://www.ncbi.nlm.nih.gov>

analyzed by SMART². The sequences of another published MADS-box genes were obtained from NCBI using BLASTP³. Multiple alignments were conducted with ClustalX 2.0 and viewed with GeneDOC. Sequences selected for the phylogenetic analysis were GmAGL1, the ABCDE classes of MADS-box proteins from *Arabidopsis* and previously reported AG homologs from several species. The accession numbers of the protein sequences are listed in Supplementary S2. The ML (Maximum Likelihood) phylogenetic tree was constructed using MEGA 6.0 with the following parameters: bootstrap method with 1,000 replications, “Jones-Taylor-Thornton (JTT)” as the substitution model, “complete deletion” for gaps/missing data treatment and other parameters with a default value.

Gene Expression Analysis

Total RNA was isolated using an RNA Plant Extraction Kit (TIANGEN, China) and reverse transcribed by AMV reverse transcriptase (Takara, Japan) as described in the manufacturer's instructions. *GmAGL1* specific primers were synthesized as follows: sense: 5'-GCTGAACACGAGAAAGCACA-3'; anti-sense: 5'-GGCACTCTCCTTCACGAAAC-3'. Semi-quantitative RT-PCR assay was performed as previously described (Huang et al., 2006). Real-time PCR was carried out with a Bio-RAD iQ5 real-time PCR system (Bio-Rad, USA) using SYBR® Green Real-time PCR Master Mix QPK-201 (Toyobo Japan). The Tubulin gene (GenBank accession No: AY907703) was quantified to normalize the amount of total transcript. The relative expression of *GmAGL1* was calculated according to the method of $2^{-\Delta\Delta Ct}$ (Livak and Schmittgen, 2001).

For the expression analysis of MADS-box genes in transgenic *Arabidopsis*, total RNA was prepared from the flowers of the wild-type and 35S:GmAGL1 plants. The *Arabidopsis* TIP4 gene was used as an internal control. The primers for each MADS-box gene are shown in Supplementary Table S2.

Subcellular Localization of GmAGL1

The coding region of *GmAGL1* was amplified with a sense primer (5'-CTAGTCTAGAATGGAATTTCCCAACGAAGC-3') containing an Xba I site and an anti-sense primer (5'-CGCGGATCCGACAAGTTGAAGAGCAGTCTGGTC-3') containing a BamH I site. The PCR product was correctly inserted into the Xba I and BamH I sites of vector pAN580, generating a GmAGL1:GFP in-frame fusion protein. This construct was then introduced into *Arabidopsis* leaf protoplasts via PEG-mediated transformation. The protoplasts were incubated under weak light for 12 h to 16 h and observed with a LSM 700 exciter confocal laser scanning microscope (Carl Zeiss). The excitation wavelength used for GFP was 488 nm.

Ectopic Expression in Arabidopsis

The plant overexpression vector pMDC32 was used for *Arabidopsis* transformation and contained a double 35S CaMV promoter. The *GmAGL1* ORF from the start to stop codons was cloned into pMDC32 by Gateway™ Technology (Invitrogen

Shanghai). Col-0 plants were transformed with this construct by *Agrobacterium tumefaciens* using the floral dip method. Transgenic seeds were germinated on solid MS medium containing 20 µg/mL Hygromycin B. Resistant seedlings were transferred to soil and further analyzed by PCR and qRT-PCR.

Scanning Electron Microscopy

Fruits from wild-type and 35S:GmAGL1 *Arabidopsis* plants were vacuum infiltrated with 4% glutaraldehyde in 25 mM phosphate buffer (pH 7.0) for 10 min and fixed with fresh solution for 16 h at 4°C, washed subsequently in 25 mM phosphate buffer (pH 7.0) and incubated for 4 h in 1% osmic acid. Samples were dehydrated gradually in an ethanol series of 30, 50, 70, 80, 95, and 100% and then critical point dried in liquid CO₂. Dried samples were covered with gold in a Nanotech sputter coater and examined with a scanning electron microscope (Philips SEM-505).

Yeast Two-Hybrid Assay

To screen the GmAGL1 interaction proteins *in vivo*, the yeast two-hybrid (Y2H) screen of the soybean cDNA library was performed by the ProQuest™ Two-Hybrid System (Invitrogen). The coding region of *GmAGL1* was recombined in a pDEST32 vector that carried a GAL4 DNA-binding domain to generate the bait construct. The Y2H cDNA library prepared from soybean flowers was used as prey and co-transformed with the bait into MaV203 yeast competent cells according to the manufacturer's protocol. Transformants were then cultured on SC-Leu-Trp-His-master plates supplemented with 40 mM of 3-aminotriazole (3-AT). The positive clones were verified by retransformation.

Bimolecular Fluorescence Complementation (BiFC)

The coding sequences for GmAGL1, GmSEP1 (GenBank accession No: DQ159905) (Huang et al., 2009) and GmSEP3 (GenBank accession No: AJ878424) (Huang et al., 2014), were cloned into vector pUC-SPYCE (C-YFP, 156–239 amino acid) to generate the C-terminal in-frame fusions with C-YFP, whereas GmAGL1 coding sequences were introduced into pUC-SPYNE (N-YFP, 1–155 amino acid) to form N-terminal in-frame fusions with N-YFP. The plasmids were co-transformed into onion epidermal cells to verify the interaction between proteins by gold particle bombardment, with a concentration ratio of 1: 1. Two days after bombardment, imaging of co-expressed YFP fragment signals was examined with a confocal fluorescence microscope. The primers used for bimolecular fluorescence complementation (BiFC) are shown in Supplementary Table S3.

RESULTS

Isolation and Sequence Analysis of AGAMOUS-LIKE1 from Soybean

The complete *GmAGL1* including 729 bp of ORF (GenBank accession No: KY321171) was isolated from soybean flower cDNA by RT-PCR and sequenced (Supplementary S1). *GmAGL1* encodes a predicted protein of 242 amino acids with a calculated

²<http://smart.embl-heidelberg.de/>

³<http://www.ncbi.nlm.nih.gov/blast>

molecular mass of 27.90 kDa and a pI of 9.55. GmAGL1 contained the conserved domains that characterize MADS-box proteins: MADS domain, I domain and K domain (**Figure 1**). Although the C-terminal region was highly divergent, there were two short conserved motifs: the AG motifs I and II (**Figure 1**). Alignment analysis of amino acid sequences showed that GmAGL1 shared 86% identity with M8 (*Pisum sativum*, AAX69070) and 84% identity with LjAGL1 (*Lotus japonicus*, AAX13305). Compared with other well-known MADS-box

proteins, GmAGL1 was 72% identical to SHP1/2 of *Arabidopsis* (Liljegren et al., 2000) and 78% identical to PERSHP of Peach (Tani et al., 2007).

To determine the evolutionary relationship of GmAGL1 with other known ABCDE class proteins from *Arabidopsis*, we used the overall amino acid sequences were used for phylogenetic analysis. GmAGL1 is located in the AGAMOUS (AG) subfamily that comprises the conserved euAG and PLENA (PLE) lineages in core eudicots (**Figure 1**) (Kramer et al., 2004). GmAGL1 was highly homologous to PLE proteins in legumes and *Arabidopsis*, suggesting that it belonged to PLE lineage and may function as a class D gene in the ABCDE model.

Expression Patterns of the GmAGL1 during Reproductive Growth

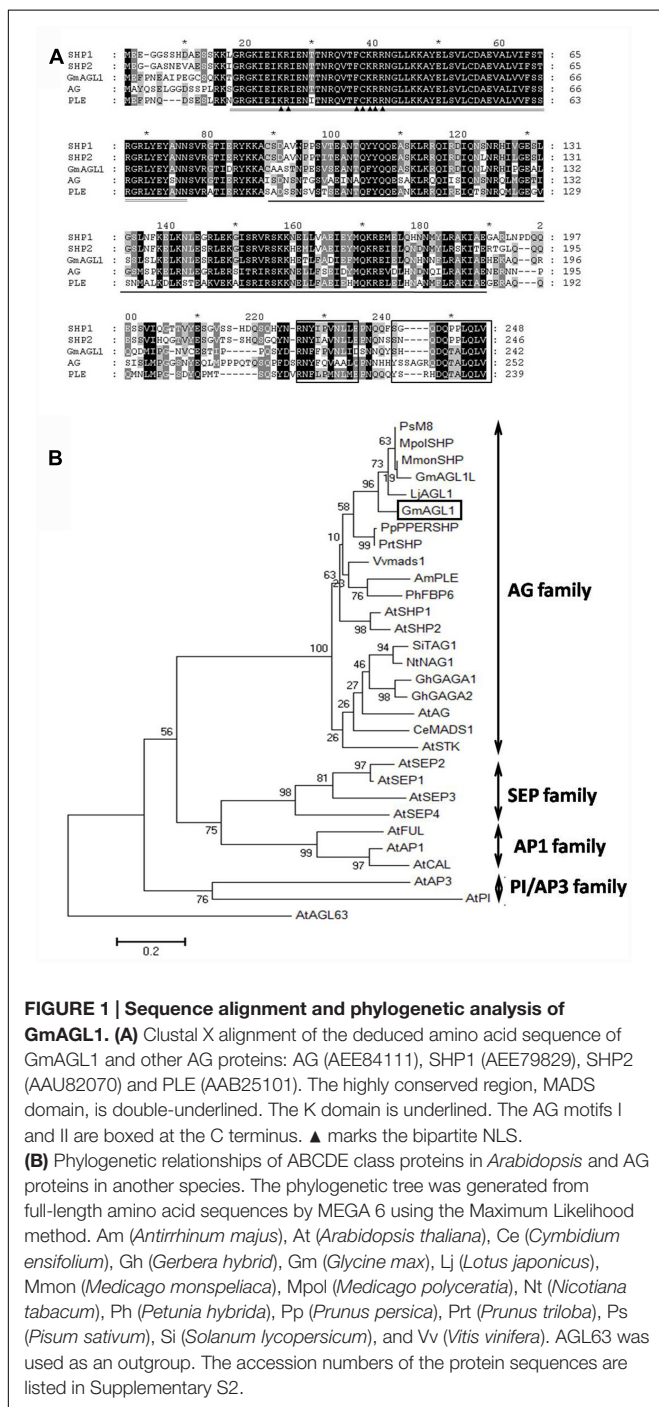
GmAGL1 is specifically expressed in flowers and pods of different stages (**Figures 2C–F**), but not in roots, stems, and leaves (**Figure 2A**). Its expression was detected at early stages and increased gradually with the process of flower and pod development (**Figures 2C–F**). In floral organs, *GmAGL1* transcripts showed strong expression in carpels and less expression in sepals (**Figure 2B**). In the developing seed, *GmAGL1* was exclusively expressed in the seed coat, which was derived from ovule integument after fertilization (Supplementary Figure S1). The gene expression profile indicates that *GmAGL1* may play important roles in carpel and pod development.

Nuclear Localization of GmAGL1

Because the highly conserved MADS domain contains a putative nuclear localization signal (NLS) (Immink et al., 2002), GmAGL1 was predicted to be localized in the nucleus (**Figure 1**). To confirm the localization in plant cells, GmAGL1 was fused with GFP, under the control of a 35S promoter, and transiently expressed in *Arabidopsis* protoplasts. Confocal green fluorescence imaging revealed that GmAGL1–GFP fusion protein was located in the nucleus, whereas free GFPs, 35S:GFP, were distributed throughout the whole cell (**Figure 3**).

Phenotypic Changes in GmAGL1 Ectopic Expression Lines

To characterize the function of GmAGL1 further, we examined transgenic *Arabidopsis* plants that were constitutively expressed *GmAGL1* under the control of the cauliflower mosaic virus 35S promoter (35S:GmAGL1). The overexpression of *GmAGL1* significantly affected the development of the transgenic *Arabidopsis*, which flowered substantially earlier than the flowers of wild-type plants (**Figure 4A**). The number of leaves produced in the transgenic lines at bolting was significantly different from that in the wild-type plants. The transgenic lines produced on average 9 leaves before flowering, whereas wild-type plants produced approximately 17 leaves (**Figure 4B**). Flowers from plants highly expressing GmAGL1 showed an abnormal morphology, in which all the petals were absent (**Figures 5A–D**). After fertilization, the senescence and abscission of sepals were significantly delayed in transgenic plants, and green



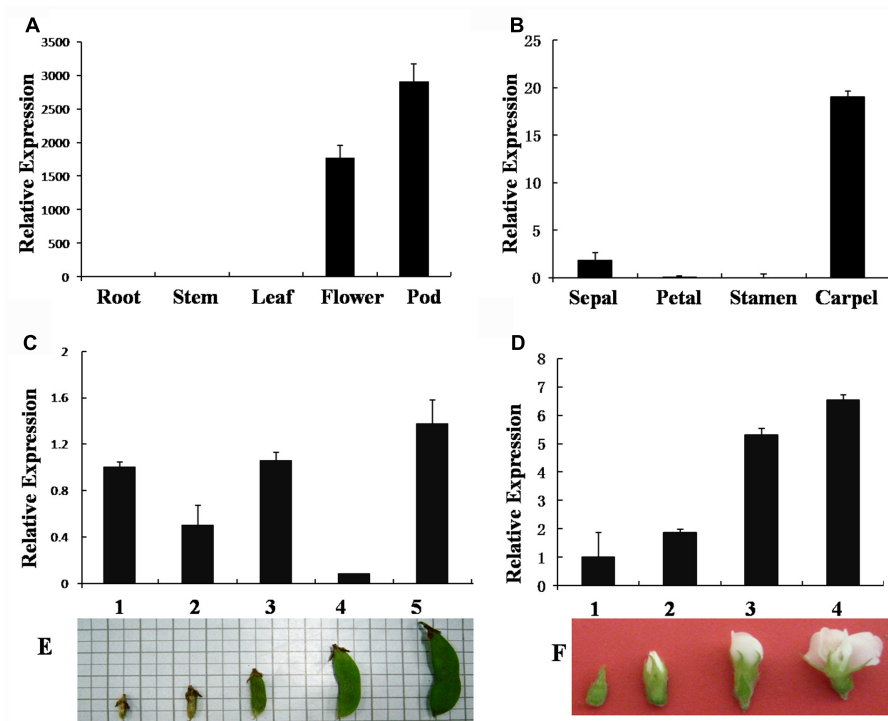


FIGURE 2 | Real-time PCR analysis of *GmAGL1* expression in different organs and at different developmental stages. (A) Gene expression profile of *GmAGL1* in vegetative organs, flowers and pods. **(B)** Expression profile of *GmAGL1* in floral organs. **(C)** Expression profile of *GmAGL1* in developing pods at the five developmental stages (1–5) illustrated in **(E)**. **(D)** Expression profile of *GmAGL1* in developing flowers at the four developmental stages (1–4) illustrated in **(F)**. The error bars represent SD based on three replicates.

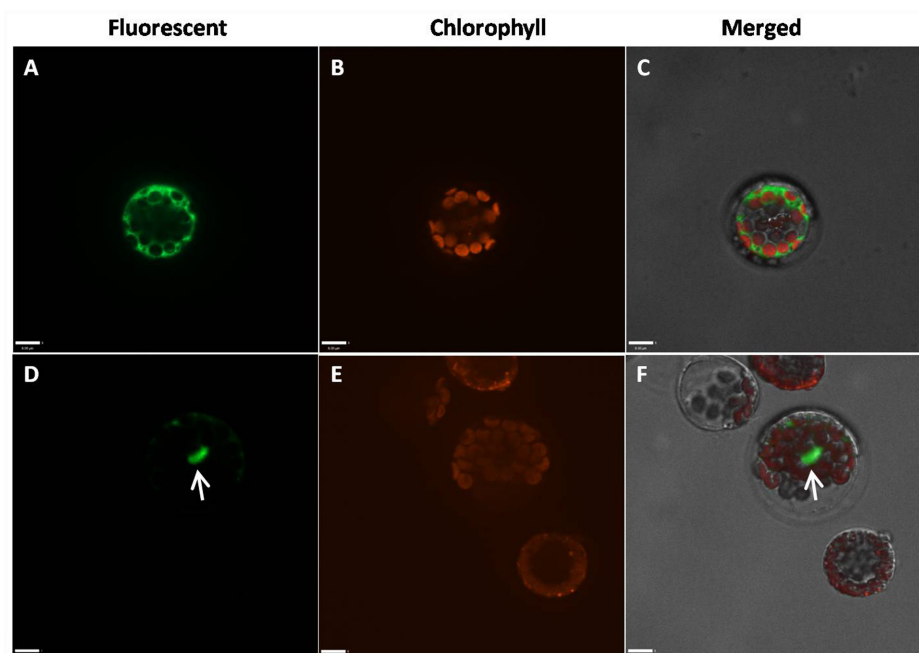


FIGURE 3 | Subcellular localization of *GmAGL1*. *GmAGL1*-GFP fusion proteins were driven by CaMV 35S promoter and transiently expressed in *Arabidopsis* leaf protoplasts. Photographs were obtained with a confocal microscope. **(A–C)** are images of 35S:GFP as controls; **(D–F)** are images of *GmAGL1*-GFP. The arrow shows the location of *GmAGL1*. Scale bars = 8 μ m.

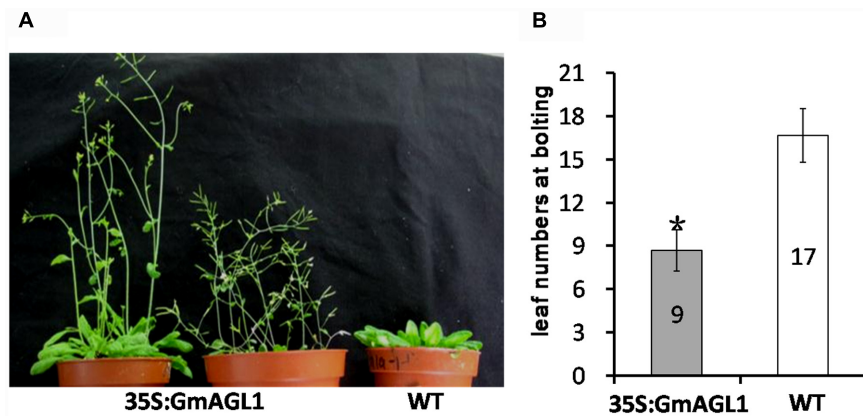


FIGURE 4 | Early flowering in 35S:GmAGL1 transgenic *Arabidopsis* plants. (A) Image of wild-type plants (WT) and transgenic plants (35S:GmAGL1) overexpressing GmAGL1 at 6 weeks after germination. **(B)** Comparison of leaf numbers at bolting between 35S: GmAGL1 and WT plants. Values correspond to the average leaf numbers at bolting. Error bars represent the standard deviation. An asterisk indicates a significant difference (*t*-test, $P < 0.01$) between transgenic and wild-type plants.

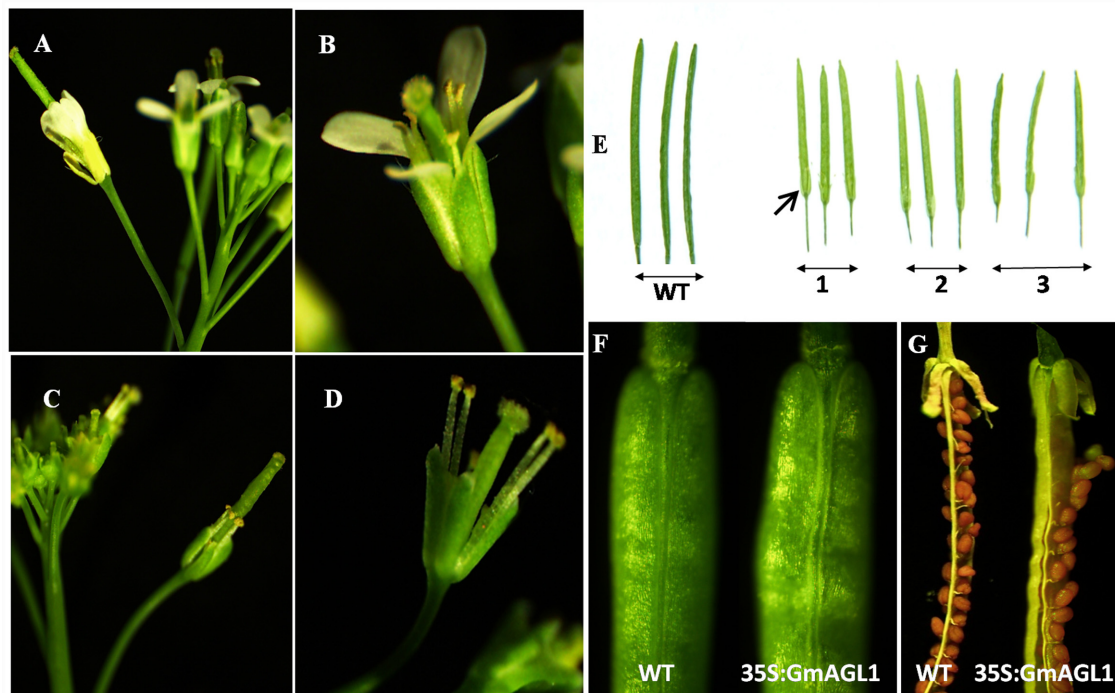


FIGURE 5 | Abnormal morphology of flowers and siliques in 35S:GmAGL1 transgenic *Arabidopsis* plants. (A,B) The wild-type flowers at anthesis showing four petals surrounding the stamen and stigma. **(C,D)** 35S:GmAGL1 flowers. Note the complete absence of petals. **(E)** The siliques from three 35S:GmAGL1 lines (1–3) are short and yellowish-green with unabscised sepals, compared to those from wild-type (WT). **(F)** The closer view of the siliques. **(G)** The silique of 35S:GmAGL1 dehiscence before the valves turned yellow and the seeds reach full maturity, which was much earlier than in wild-type (WT).

unabscised sepals were observed even in the matured siliques (Figure 5E).

Moreover, phenotypic differences between the developing fruits of 35S:GmAGL1 and wild-type plants were also observed. 35S:GmAGL1 fruits were shorter in length and appeared yellowish-green compared to the long and dark green color of wild-type fruits (Figure 5E). The valve margins of 35S:GmAGL1

fruits were more remarkable than those of wild-type fruits (Figure 5F). To gain a clear observation, scanning electron microscopy was used to examine the morphological changes in the fruits. The outer epidermis of 35S:GmAGL1 fruit valves differed from that of wild-type fruit valves. The valve margins were significantly wider and thinner compared to wild-type fruits (Figures 6A–D). The microscopic cross section of the transgenic

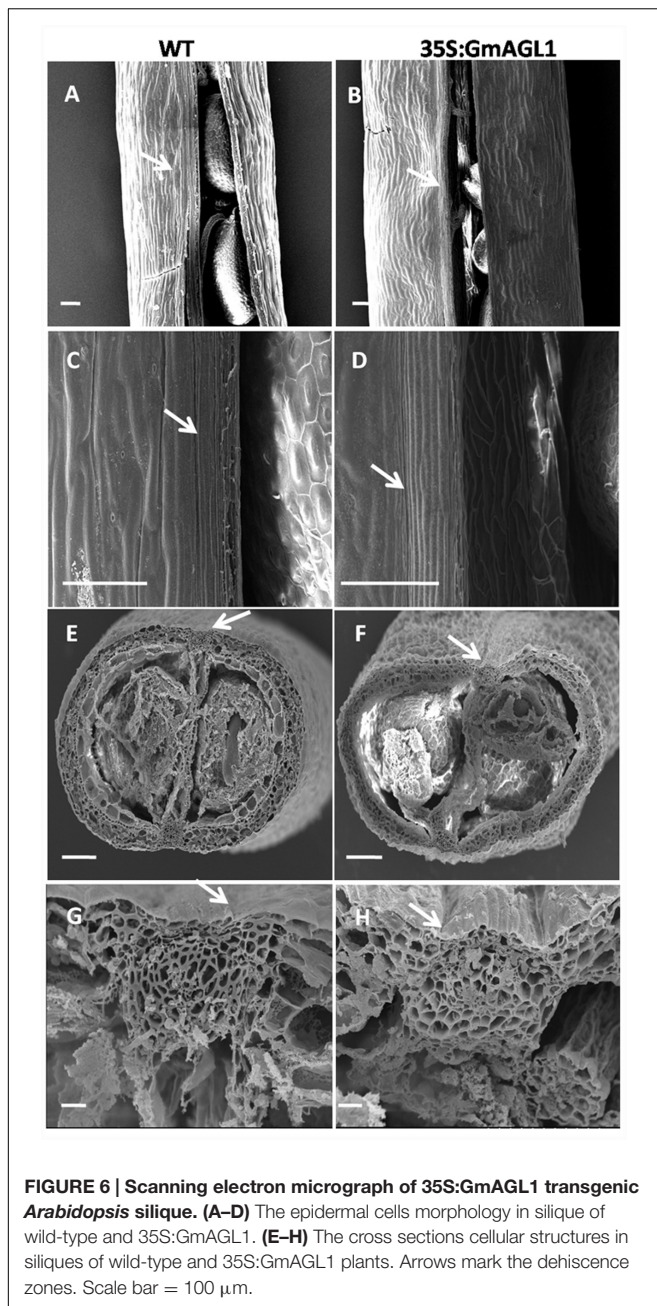


FIGURE 6 | Scanning electron micrograph of 35S:GmAGL1 transgenic *Arabidopsis* silique. (A–D) The epidermal cells morphology in silique of wild-type and 35S:GmAGL1. **(E–H)** The cross sections cellular structures in siliques of wild-type and 35S:GmAGL1 plants. Arrows mark the dehiscence zones. Scale bar = 100 μ m.

siliques cracking are showed a less condensed structure and reduced thickness (**Figures 6E–H**). In the premature fruits, dehiscence occurred in the valve regions while the seeds were still fresh and green, resulting in early seed dispersal (**Figure 5G**). These gain-of-function analyses suggest that GmAGL1 has a direct role in promoting valve development and regulating fruit elongation and dehiscence.

Expression Analysis of Genes in Transgenic *Arabidopsis*

GmAGL1 was localized to the nuclear compartments of plant cells, indicating that it might regulate gene expression as a

transcription factor. Based on the function of GmAGL1, as revealed by its constitutive expression, we studied the expression of five genes involved in flowering and fruit development in the 35S:GmAGL1 flowers. It was observed that the *Arabidopsis* homologs of the fruit dehiscence related genes, *ALC*, *IND* and *STK*, showed more accumulation in 35S:GmAGL1 compared to wild-type plants (**Figure 7**). As will be described later in the Y2H screen, GmAGL1 interacted with SEP homologs. The expression levels of both *SEP1* and *SEP3* were increased in the 35S:GmAGL1 *Arabidopsis* (**Figure 7**). These results suggested that GmAGL1 might directly regulate the expression of these genes to control flower development and fruit dehiscence.

Screening of GmAGL1 Interaction Proteins in Flowers

We applied a GAL4-based experiment to screen for the proteins that interact directly with GmAGL1. As shown in Supplementary Figure S2, the yeast cells containing BD-GmAGL1 could not grow on the SC-Leu-Try-His+40 mM 3AT medium, nor could the negative self-activation control. This suggests that GmAGL1 exhibits no transactivation activity and can be used as bait to perform a Y2H screen. As a result, we identified four proteins as interacting proteins of GmAGL1, which were named Interaction Protein 1 to Interaction Protein 4 (IP1~4) (Supplementary Table S1). Two of these were SEP1-like proteins (IP1/3), one was a SEP3-like protein (IP2), and one was a putative CHUP1 protein (IP4). IP1 (Glyma18g50900), IP2 (Glyma08g11120), and IP3 (Glyma08g27670) were highly similar to homologs of SEP in the MADS-box gene family. This result suggests that GmAGL1 and SEP-like proteins in soybean can interact in specific manners and form macromolecular complexes to regulate flower and fruit development.

GmAGL1 Interacts with Soybean SEP Homologs in the Nucleus

The interactions between GmAGL1 and SEP homologs were further confirmed *in vivo* by BiFC assay. Two SEP functional homologs studied in soybean (Huang et al., 2009, 2014) were inserted into pUC-SPYCE (GmSEP1/3-C-YFP). Using particle bombardment, GmAGL1-N-YFP and GmSEP1/3-C-YFP were co-transformed into onion epidermal cells. Fluorescence confocal microscopy showed that BiFC signals were present in the nuclear compartments of transformed cells (**Figure 8**). GmAGL1-N-YFP and pUC-SPYCE, pUC-SPYNE and GmSEP1/3-C-YFP, pUC-SPYNE and pUC-SPYCE – as negative controls – were separately bombarded into onion epidermal cells, after which no fluorescence was detected (Supplementary Figure S3). These experiments demonstrated that GmAGL1 interacted with soybean SEP homologs in the nucleus.

GmAGL1 Cannot Form Homodimers

The BIFC system was also used to detect whether GmAGL1 can form homodimers. Full-length GmAGL1 sequences were inserted into pUC-SPYNE and pUC-SPYCE to construct the recombinant plasmids GmAGL1-N-YFP and GmAGL1-C-YFP, respectively, where GmAGL1 was in-frame fused with the

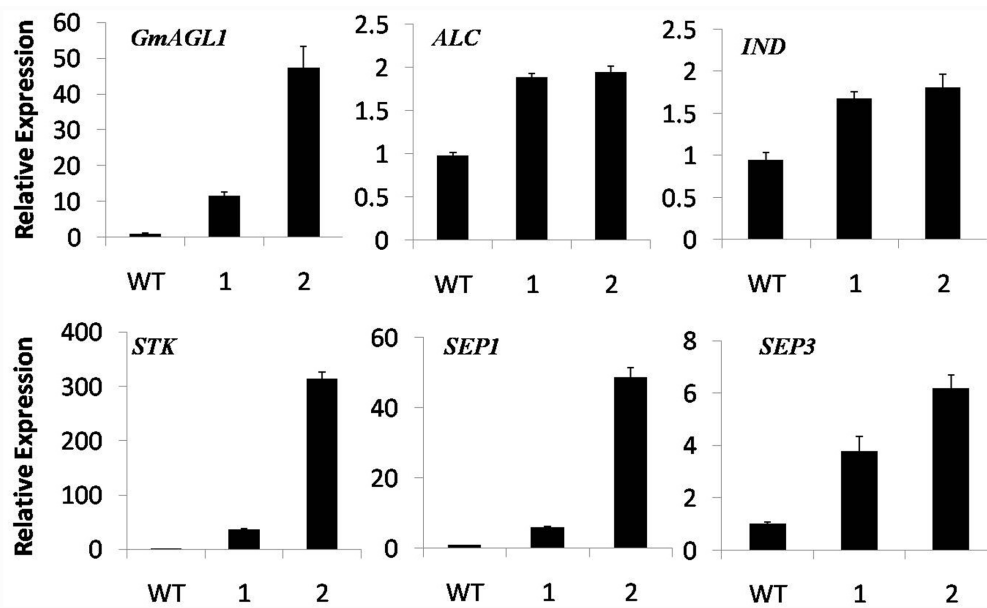


FIGURE 7 | Real-time PCR analysis of flowering and fruit dehiscence related gene expression in 35S:GmAGL1 transgenic *Arabidopsis* flowers. The TIP4 gene was used as an internal control. The primers for quantitative RT-PCR are shown in Supplementary Table S2. The error bars represent SD based on three replicates.

N-terminus and C-terminus of YFP. GmAGL1-N-YFP and GmAGL1-C-YFP were co-expressed in onion epidermal cells through particle bombardment, but no YFP signals were found (Figure 8). These results indicated that GmAGL1 did not interact with itself in plant cells, i.e., GmAGL1 cannot form homodimers.

DISCUSSION

The MADS-box transcription factor family is a large family that controls flower and fruit patterning in plants. To date, many members of this family have been identified and extensively studied in different plant species. However, much less is known about their functions in the developmental processes of soybean (Huang et al., 2014). In a previous study, genome wide expression profiles of soybean genes were investigated by Affymetrix Gene Chip in roots, leaves, flowers and pods (Huang et al., 2009). Some MADS-box genes were found to be primarily expressed in both flowers and pods. The roles of these genes are still not clear and require further functional analysis. In this study, we functionally characterized a soybean MADS-box gene, *GmAGL1*.

GmAGL1 is a MADS-box type MADS-box protein encoded by genomic DNA, as it possesses a modular structure where the MADS (M) domain is followed by an intervening (I), a keratin (K) and a C-terminal (C) domain (Theissen et al., 1996; Kaufmann et al., 2005). Two short regions of high conservation, the AG motifs I and II, were identified in the variable C domain. These two motifs are conserved in the AG subfamily (Kramer et al., 2004). Ectopic expression of an AG protein lacking the C-terminal produced *ag*-like flowers in transgenic *Arabidopsis*, indicating that these AG motifs are required for the

correct function of AG proteins in plant development (Mizukami et al., 1996). Phylogenetic analyses also showed that GmAGL1 was clustered into the AG subfamily and closely allied with the PLE lineage. Members of the AG subfamily have been characterized from angiosperms as master regulators of stamen, carpel and ovule identities. They play important roles after fertilization in the developing fruits and seeds. These studies also demonstrated that AG-like genes retain functional conservation within flowering plants (Dreni and Kater, 2014). The functional conservation of AG homologs suggests that *GmAGL1* might have central roles in regulating soybean floral organ identity and pod development.

Previous results showed that *GmAGL1* was exclusively expressed in flowers and pods. We found that *GmAGL1* was specifically expressed in carpels, pod walls and seed coats, but only weak expression was seen in sepals. The transcript was detected at all stages of flower and pod development. The expression pattern of *GmAGL1* was similar to that observed for other AG-like genes. For instance, the transcript of *PPERSHP* was detected primarily in carpels and accumulated throughout fruit development from full anthesis until fruit harvest (Tani et al., 2007). *AGL1* was preferentially expressed in particular regions of the gynoecium and ovule, only during and after floral development (Flanagan et al., 1996). The spatial and temporal expression profiles of AG-like genes are closely aligned with their conserved functions in carpel and pod development.

Consistent with a role as a transcription factor, GmAGL1 was localized in the nucleus and contained putative DNA-binding domain in the N-terminus. Nevertheless, none of the putative motifs related to transcriptional activation were found in the conserved domain analysis of GmAGL1. The full length of

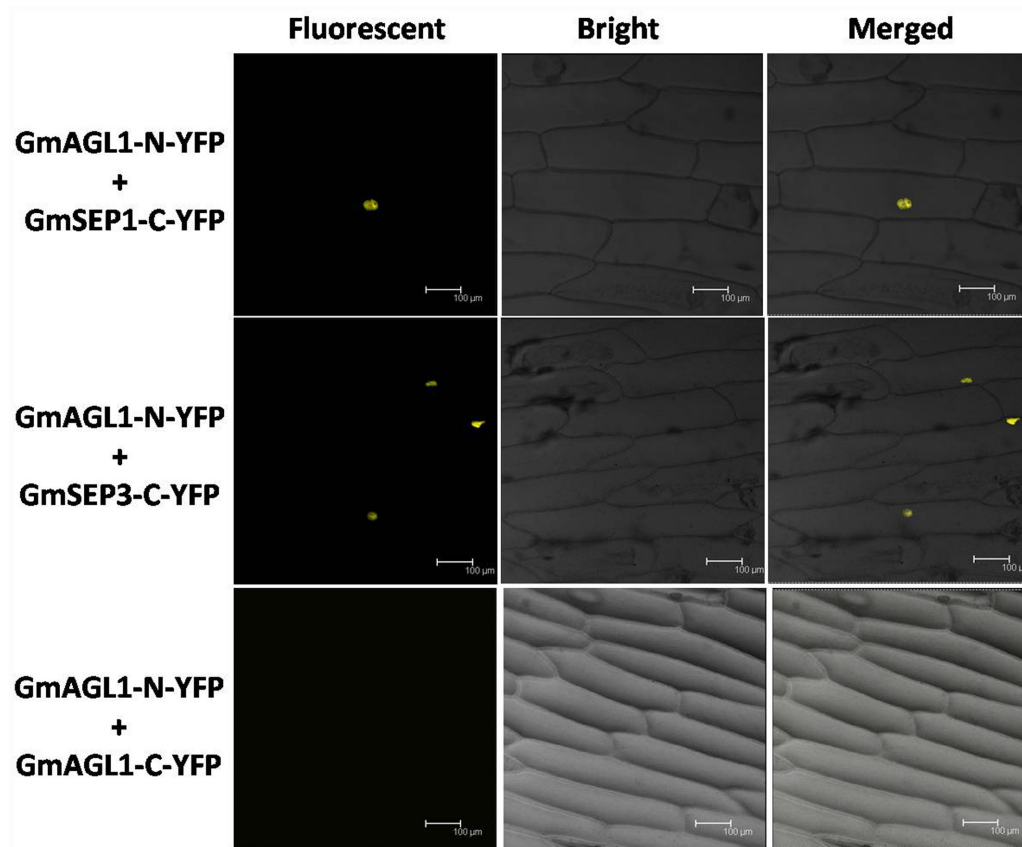


FIGURE 8 | Bimolecular fluorescence complementation (BiFC) analysis of GmAGL1 interactions *in planta*. Fluorescence was observed in the nuclear compartment of onion epidermal cells and results from complementation of the N-terminal part of the YFP fused with GmAGL1 (GmAGL1-N-YFP) with the C-terminal part of the YFP fused with soybean SEP homologs (GmSEP1/3-C-YFP). No fluorescence was observed when GmAGL1-N-YFP was co-expressed with GmAGL1-C-YFP. Bright-field images, YFP epifluorescence images and overlay images of the same cells are shown. Scale bar = 100 μ m.

GmAGL1 also exhibited inactivity in the self-activation test in the Y2H system. Thus far, the ability of transactivation has not been described in AGL1 proteins. Interestingly, the proteins that directly interact with GmAGL1 were obtained from a soybean flower cDNA library by a yeast two-hybrid screen. It was also reported that AP1 and SEP3 added transcriptional activator activity to PI-AP3 and AG by the formation of complexes (Honma and Goto, 2001). Therefore, we predict that GmAGL1 might allow transcriptional regulation of target genes as an activator in a combinatorial way by interacting with other factors.

It is well-known that MADS domain proteins do not exert their functions as monomers, but rather they form multimeric protein complexes with other proteins (de Folter et al., 2005; Immink et al., 2010). In this study, a screen of a cDNA expression library with the yeast two-hybrid GAL4 system was performed to unravel the protein-protein interaction network for GmAGL1. As a result, three putative MADS-box transcription factors were identified as GmAGL1 interaction partners from soybean flowers. Interestingly, all these proteins belonged to the SEP subfamily, indicating that the function of GmAGL1 in promoting carpel identity may be based on a biochemical interaction with SEP proteins. In *Arabidopsis*, three SEP factors, SEP1, SEP2 and

SEP3, which are closely related and functionally redundant, were necessary to determine the identities of petals, stamens, and carpels (Pelaz et al., 2000). In addition, protein-protein interaction studies have revealed that MADS-box proteins are dependent for their function of floral organ identity on the interaction with SEP proteins (Honma and Goto, 2001). These findings contributed to the proposal of the genetic ABCDE model for flower development, which demonstrated that the members of SEP proteins (E-class) act as bridges allowing the formation of higher order complexes (Theissen and Saedler, 2001). The study also showed that the complexes composed of AGL1 and SEP homologous proteins are probably able to promote carpel identity in the absence of AG and AP2 (Favaro et al., 2003).

To investigate the biological function *in planta*, we overexpressed GmAGL1 in transgenic *Arabidopsis*. We observed flowers with abnormal morphology wherein the petals were absent. The senescence and abscission of sepals were delayed in the developing transgenic siliques. These observations suggested that GmAGL1 may have a similar activity to the C function gene AG, whose overexpression caused homeotic conversion of perianth organs by suppressing A function genes (Mizukami and Ma, 1995). In *Arabidopsis*, SHP1/2 have retained the ability

to substitute for AG activity, as the flowers constitutively expressing *SHP1* or *SHP2* showed a phenotype similar to those constitutively expressing AG (Liljegren et al., 2000). The introduction of 35S:*SHP2* into *ag* mutants was sufficient to rescue stamen and carpel development (Pinyopich et al., 2003). Many studies have shown that the overexpression of AG-like genes would alter floral organ identities in the two perianth whorls (Liljegren et al., 2000; Boss et al., 2001; Vrebalov et al., 2009; Gimenez et al., 2010). In addition, GmAGL1 interacted with soybean SEP-like proteins and promoted *SEP* expression in *Arabidopsis*. SEP-like proteins are necessary to determine the identities of petals, stamens and carpels (Pelaz et al., 2000; Huang et al., 2014). GmAGL1 might interfere with the activity of petal specification genes through competition for interacting partners.

In general, fruits are derived directly from the carpels. Therefore, any mutation that affects carpel development has an effect on fruit development. *GmAGL1* is also important for fruit development. The overexpression of *GmAGL1* in *Arabidopsis*, which has a dry dehiscent fruit similar to the soybean pod, resulted in striking phenotypic effects in the 35S:*GmAGL1* lines fruits, which were short, yellowish-green and early dehiscent. Microscopy revealed that valve margins were more visibly constricted in the developing transgenic fruits, which led to the dehiscence of valves before the seeds reach full maturity. In *Arabidopsis*, there has been significant progress in the molecular mechanisms of fruit dehiscence. *ALC* and *IND* encode two bHLH transcription factors related to *Arabidopsis* pod shatter, which is regulated by SHP in the genetic regulatory hierarchy (Rajani and Sundaresan, 2001; Liljegren et al., 2004). *STK* is a D-type MADS-box gene that is required for normal development of the funiculus, an umbilical-cord-like structure that connects the developing seed to the fruit, and for dispersal of the seeds when the fruit matures (Pinyopich et al., 2003). As expected, the overexpression of *GmAGL1* activated the expression of native *ALC*, *IND* and *STK*, which might be the molecular mechanism for the phenotypic effects of transgenic *Arabidopsis*. The functional studies of other AG-like genes gave very similar results, not only in *Arabidopsis* (Liljegren et al., 2000) but also in tobacco (Fourquin and Ferrandiz, 2012) and tomato

(Gimenez et al., 2010), suggesting that AG-like genes may play a prominent role late in fruit development and dehiscence that is generally conserved in other eudicots. Pod dehiscence is a major cause of yield loss in mechanical harvesting of soybeans, especially in several countries in tropical and sub-tropical regions (Tiwari and Bhatia, 1995). Our results indicate that further analysis of the molecular network underlying fruit dehiscence may contribute to the potential genetic manipulation of pod shattering in soybean plants.

AUTHOR CONTRIBUTIONS

FH and DY conceived this project. YC and TW designed and performed the sequence characterization, expression profile and protein interaction of GmAGL1. GX, HY, and XZ conducted the plant transformation and microscopy analysis. YC and FH wrote the article. YS and DY contributed to scientific discussions and critical revision of manuscript. All authors reviewed the final manuscript.

FUNDING

This work is supported in part by the National Basic Research Program of China project from the Ministry of Agriculture of China for Transgenic Research (2015ZX08004-003), the National Basic Research Program of China (973 Program) (2010CB125906), the National Natural Science Foundation of China (31371644, 31571688, 31601324), and the Jiangsu Collaborative Innovation Center for Modern Crop Production (JCIC-MCP).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00175/full#supplementary-material>

REFERENCES

- Arora, R., Agarwal, P., Ray, S., Singh, A. K., Singh, V. P., Tyagi, A. K., et al. (2007). MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* 8:242. doi: 10.1186/1471-2164-8-242
- Boss, P. K., Vivier, M., Matsumoto, S., Dry, I. B., and Thomas, M. R. (2001). A cDNA from grapevine (*Vitis vinifera* L.), which shows homology to AGAMOUS and SHATTERPROOF, is not only expressed in flowers but also throughout berry development. *Plant Mol. Biol.* 45, 541–553. doi: 10.1023/A:1010634132156
- Colombo, L., Franken, J., Koetje, E., van Went, J., Dons, H. J., Angenent, G. C., et al. (1995). The *Petunia* mads box gene Fbp11 determines ovule identity. *Plant Cell* 7, 1859–1868. doi: 10.2307/3870193
- de Folter, S., Immink, R. G., Kieffer, M., Parenicová, L., Henz, S. R., Weigel, D., et al. (2005). Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *Plant Cell* 17, 1424–1433. doi: 10.1105/tpc.105.031831
- Dreni, L., and Kater, M. M. (2014). MADS reloaded: evolution of the AGAMOUS subfamily genes. *New Phytol.* 201, 717–732. doi: 10.1111/nph.12555
- Favaro, R., Pinyopich, A., Battaglia, R., Kooiker, M., Borghi, L., Ditta, G., et al. (2003). MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* 15, 2603–2611. doi: 10.1105/tpc.015123
- Ferrandiz, C., Liljegren, S. J., and Yanofsky, M. F. (2000). Negative regulation of the SHATTERPROOF genes by FRUITFULL during *Arabidopsis* fruit development. *Science* 289, 436–438. doi: 10.1126/science.289.5478.436
- Flanagan, C. A., Hu, Y., and Ma, H. (1996). Specific expression of the AGL1 MADS-box gene suggests regulatory functions in *Arabidopsis* gynoecium and ovule development. *Plant J.* 10, 343–353. doi: 10.1046/j.1365-313X.1996.1002.0343.x
- Fourquin, C., and Ferrandiz, C. (2012). Functional analyses of AGAMOUS family members in *Nicotiana benthamiana* clarify the evolution of early and late roles of C-function genes in eudicots. *Plant J.* 71, 990–1001. doi: 10.1111/j.1365-313X.2012.05046.x
- Gimenez, E., Pineda, B., Capel, J., Antón, M. T., Atarés, A., Pérez-Martín, F., et al. (2010). Functional analysis of the arlequin mutant corroborates the essential role of the ARLEQUIN/TAGL1 gene during reproductive development of tomato. *PLoS ONE* 5:e14427. doi: 10.1371/journal.pone.0014427

- Goto, K., and Meyerowitz, E. M. (1994). Function and regulation of the *Arabidopsis* floral homeotic gene PISTILLATA. *Genes Dev.* 8, 1548–1560. doi: 10.1101/gad.8.13.1548
- Honma, T., and Goto, K. (2001). Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature* 409, 525–529. doi: 10.1038/35054083
- Huang, F., Chi, Y., Gai, J., and Yu, D. (2009). Identification of transcription factors predominantly expressed in soybean flowers and characterization of GmSEP1 encoding a SEPALLATA1-like protein. *Gene* 438, 40–48. doi: 10.1016/j.gene.2009.03.002
- Huang, F., Chi, Y., Meng, Q., Gai, J., and Yu, D. (2006). GmZFP1 encoding a single zinc finger protein is expressed with enhancement in reproductive organs and late seed development in soybean (*Glycine max*). *Mol. Biol. Rep.* 33, 279–285. doi: 10.1007/s11033-006-9012-z
- Huang, F., Xu, G., Chi, Y., Liu, H., Xue, Q., and Zhao, T. (2014). A soybean MADS-box protein modulates floral organ numbers, petal identity and sterility. *BMC Plant Biol.* 14:89. doi: 10.1186/1471-2229-14-89
- Immink, R. G., Gadella, T. W. Jr., Ferrario, S., Busscher, M., and Angenent, G. C. (2002). Analysis of MADS box protein-protein interactions in living plant cells. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2416–2421. doi: 10.1073/pnas.042677699
- Immink, R. G., Kaufmann, K., and Angenent, G. C. (2010). The ‘ABC’ of MADS domain protein behaviour and interactions. *Semin. Cell Dev. Biol.* 21, 87–93. doi: 10.1016/j.semdb.2009.10.004
- Jack, T., Brockman, L. L., and Meyerowitz, E. M. (1992). The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell* 68, 683–697. doi: 10.1016/0092-8674(92)90144-2
- Jofuku, K. D., den Boer, B. G., Van Montagu, M., and Okamoto, J. K. (1994). Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6, 1211–1225. doi: 10.1105/tpc.6.9.1211
- Karlova, R., Chapman, N., David, K., Angenent, G. C., Seymour, G. B., and de Maagd, R. A. (2014). Transcriptional control of fleshy fruit development and ripening. *J. Exp. Bot.* 65, 4527–4541. doi: 10.1093/jxb/eru316
- Katahata, S. I., Futamura, N., Igasaki, T., and Shinohara, K. (2014). Functional analysis of SOC1-like and AGL6-like MADS-box genes of the gymnosperm *Cryptomeria japonica*. *Tree Genet. Genomes* 10, 317–327. doi: 10.1007/s11295-013-0686-9
- Kaufmann, K., Melzer, R., and Theissen, G. (2005). MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants. *Gene* 347, 183–198. doi: 10.1016/j.gene.2004.12.014
- Kramer, E. M., Jaramillo, M. A., and Di Stilio, V. S. (2004). Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* 166, 1011–1023. doi: 10.1534/genetics.166.2.1011
- Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089
- Liljegren, S. J., Roeder, A. H., Kempin, S. A., Gremski, K., Østergaard, L., Guimil, S., et al. (2004). Control of fruit patterning in *Arabidopsis* by INDEHISCENT. *Cell* 116, 843–853. doi: 10.1016/S0092-8674(04)00217-X
- Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2⁻(Delta Delta C(T)) Method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Ma, H., Yanofsky, M. F., and Meyerowitz, E. M. (1991). AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev.* 5, 484–495. doi: 10.1101/gad.5.3.484
- Mandel, M. A., Gustafson-Brown, C., Savidge, B., and Yanofsky, M. F. (1992). Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. *Nature* 360, 273–277. doi: 10.1038/360273a0
- Mizukami, Y., Huang, H., Tudor, M., Hu, Y., and Ma, H. (1996). Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. *Plant Cell* 8, 831–845. doi: 10.1105/tpc.8.5.831
- Mizukami, Y., and Ma, H. (1995). Separation of Ag function in floral meristem determinacy from that in reproductive organ identity by expressing antisense Ag RNA. *Plant Mol. Biol.* 28, 767–784. doi: 10.1007/BF00042064
- Pelaz, S., Ditta, G. S., Baumann, E., Wisman, E., and Yanofsky, M. F. (2000). B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 405, 200–203. doi: 10.1038/35012103
- Pinyopich, A., Ditta, G. S., Savidge, B., Liljegren, S. J., Baumann, E., Wisman, E., et al. (2003). Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* 424, 85–88. doi: 10.1038/nature01741
- Rajani, S., and Sundaresan, V. (2001). The *Arabidopsis* myc/bHLH gene ALCATRAZ enables cell separation in fruit dehiscence. *Curr. Biol.* 11, 1914–1922. doi: 10.1016/S0960-9822(01)00593-0
- Robles, P., and Pelaz, S. (2005). “Flower and fruit development in *Arabidopsis thaliana*.” *Int. J. Dev. Biol.* 49, 633–643. doi: 10.1387/ijdb.052020pr
- Tani, E., Polidoros, A. N., and Tsafaris, A. S. (2007). Characterization and expression analysis of FRUITFULL- and SHATTERPROOF-like genes from peach (*Prunus persica*) and their role in split-pit formation. *Tree Physiol.* 27, 649–659. doi: 10.1093/treephys/27.5.649
- Theissen, G., Kim, J. T., and Saedler, H. (1996). Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *J. Mol. Evol.* 43, 484–516. doi: 10.1007/BF02337521
- Theissen, G., and Saedler, H. (2001). Plant biology. Floral quartets. *Nature* 409, 469–471. doi: 10.1038/35054172
- Tiwari, S. P., and Bhatia, V. S. (1995). Characters of pod anatomy associated with resistance to pod-shattering in soybean. *Ann. Bot.* 76, 483–485. doi: 10.1006/anbo.1995.1123
- Vrebalov, J., Pan, I. L., Arroyo, A. J. M., McQuinn, R., Chung, M., Poole, M., et al. (2009). Fleshy fruit expansion and ripening are regulated by the tomato SHATTERPROOF gene TAGL1. *Plant Cell* 21, 3041–3062. doi: 10.1105/tpc.109.066936
- White, P. J. (2002). Recent advances in fruit development and ripening: an overview. *J. Exp. Bot.* 53, 1995–2000. doi: 10.1093/jxb/erf105
- Xue, L. J., Zhang, J. J., and Xue, H. W. (2012). Genome-wide analysis of the complex transcriptional networks of rice developing seeds. *PLoS ONE* 7:e31081. doi: 10.1371/journal.pone.0031081
- Yanofsky, M. F., Ma, H., Bowman, J. L., Drews, G. N., Feldmann, K. A., and Meyerowitz, E. M. (1990). The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. *Nature* 346, 35–39. doi: 10.1038/346035a0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chi, Wang, Xu, Yang, Zeng, Shen, Yu and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



De novo Transcriptome Profiling of Flowers, Flower Pedicels and Pods of *Lupinus luteus* (Yellow Lupine) Reveals Complex Expression Changes during Organ Abscission

Paulina Glazinska^{1,2*}, Waldemar Wojciechowski^{1,2}, Milena Kulasek¹,
Wojciech Glinkowski¹, Katarzyna Marciniak^{1,2}, Natalia Klajn¹, Jacek Keszy¹ and
Jan Kopcewicz¹

OPEN ACCESS

Edited by:

Diego Rubiales,
Instituto de Agricultura Sostenible,
Spain

Reviewed by:

Benedetto Ruperti,
University of Padova, Italy
Juan De Dios Alché,
Consejo Superior de Investigaciones
Científicas (CSIC), Spain
Manuel Acosta,
University of Murcia, Spain

*Correspondence:

Paulina Glazinska
paulina.glazinska@umk.pl

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 30 December 2016

Accepted: 10 April 2017

Published: 02 May 2017

Citation:

Glazinska P, Wojciechowski W,
Kulasek M, Glinkowski W, Marciniak K,
Klajn N, Keszy J and Kopcewicz J
(2017) De novo Transcriptome
Profiling of Flowers, Flower Pedicels
and Pods of *Lupinus luteus* (Yellow
Lupine) Reveals Complex Expression
Changes during Organ Abscission.
Front. Plant Sci. 8:641.
doi: 10.3389/fpls.2017.00641

¹ Department of Biology and Environmental Science, Nicolaus Copernicus University, Torun, Poland, ² Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University, Torun, Poland

Yellow lupine (*Lupinus luteus* L., Taper c.), a member of the legume family (*Fabaceae* L.), has an enormous practical importance. Its excessive flower and pod abscission represents an economic drawback, as proper flower and seed formation and development is crucial for the plant's productivity. Generative organ detachment takes place at the basis of the pedicels, within a specialized group of cells collectively known as the abscission zone (AZ). During plant growth these cells become competent to respond to specific signals that trigger separation and lead to the abolition of cell wall adhesion. Little is known about the molecular network controlling the yellow lupine organ abscission. The aim of our study was to establish the divergences and similarities in transcriptional networks in the pods, flowers and flower pedicels abscised or maintained on the plant, and to identify genes playing key roles in generative organ abscission in yellow lupine. Based on *de novo* transcriptome assembly, we identified 166,473 unigenes representing 219,514 assembled unique transcripts from flowers, flower pedicels and pods undergoing abscission and from control organs. Comparison of the cDNA libraries from dropped and control organs helped in identifying 1,343, 2,933 and 1,491 differentially expressed genes (DEGs) in the flowers, flower pedicels and pods, respectively. In DEG analyses, we focused on genes involved in phytohormonal regulation, cell wall functioning and metabolic pathways. Our results indicate that auxin, ethylene and gibberellins are some of the main factors engaged in generative organ abscission. Identified 28 DEGs common for all library comparisons are involved in cell wall functioning, protein metabolism, water homeostasis and stress response. Interestingly, among the common DEGs we also found an miR169 precursor, which is the first evidence of micro RNA engaged in abscission. A KEGG pathway enrichment analysis revealed that the identified DEGs were predominantly involved in carbohydrate and amino acid

metabolism, but some other pathways were also targeted. This study represents the first comprehensive transcriptome-based characterization of organ abscission in *L. luteus* and provides a valuable data source not only for understanding the abscission signaling pathway in yellow lupine, but also for further research aimed at improving crop yields.

Keywords: yellow lupine, RNA-Seq, DEGs, flower, pod, abscission, microRNA

INTRODUCTION

Yellow lupine (*Lupinus luteus* L.), similarly to other members of the *Fabaceae* family (*Fabaceae* L.), has an enormous practical importance. Lupine seeds contain a high storage protein level, which is why it is used as feedstock for the production of high-protein animal feed. Its symbiosis with nitrogen-fixing bacteria which support its growth and development makes this plant a natural fertilizer enriching the soil with nitrogen (Prusiński, 2007). As flower and seed formation and development in crops is crucial for their productivity, flower and pod abscission becomes a factor that reduces benefits from growing lupines (Van Steveninck, 1958, 1959; Prusiński, 2007; Wilmowicz et al., 2016). On the other hand, a moderate abscission level is an agronomically desirable characteristic, since an excessive number of fruits is inversely proportional to their quality (Dokoozlian and Peacock, 2001). In order to be able to closely control the process, full knowledge of the molecular mechanisms behind generative organ development and the signaling pathways leading to organ abscission in particular plants is required.

Abscission is the process of shedding vegetative or reproductive organs by a plant in response to developmental, hormonal, and environmental cues. This process occurs at a special layer of cells called the abscission zone (AZ), and consists in cell separation enabled by hydrolytic enzymes. Plants can abscise buds, branches, petioles, leaves, flowers and fruits, while this process can be affected by environmental factors such as temperature, light quality, disease, water stress, and nutrition (Ascough et al., 2005; Estornell et al., 2013). The abscission of plant organs is associated with changes in the auxin gradient across the AZ, which is affected by ethylene (ET). It occurs when the auxin level below the AZ is higher than its concentration above that zone (Roberts et al., 2002; Meir et al., 2010). There are four key steps in abscission: (1) the establishment of the AZ, (2) the acquisition of the competence to respond to abscission signals, (3) the activation of organ abscission, and (4) the formation of a protective layer (Kim, 2014). It has been found that before and during peduncle abscission the expression of multiple regulatory genes changes (Kim et al., 2016), and that this variation affects a number of transcription factors associated with auxin and ethylene pathways (Sundaresan et al., 2016). However, it is not only auxin and ethylene that are involved in organ dropping. Recent studies on jasmonate signaling pathway mutants *coronatine insensitive1(coi1)* and *jasmonate-ZIM domain(jazz)* showed that these hormones, too, take part in regulating flower abscission in *Nicotiana attenuate* (Oh et al., 2013). In *A. thaliana*, several genes associated with the process of organ separation were identified, these being: *BLADE ON PETIOLE (BOP)* (McKim

et al., 2008), *HAESA (HAE)* (Jinn et al., 2000), *HAESA-LIKE2 (HSL2)* (Stenvik et al., 2008), *CAST AWAY (CST)*, *NEVERSHED (NEV)*, *EVERSHED (EVR)*, *INFLORESCENCE DEFICIENT IN ABSCISSION (IDA)*. The first of them is responsible for AZ formation, the second one for AZ functioning, and the last one for the final stage of organ separation. Late abscission stages are associated with the activity of many cell wall modifying enzymes, such as: polygalacturonases (PGs) (González-Carranza et al., 2002), xyloglucan endotransglucosylases/hydrolases (XTH) (Singh et al., 2011), β -1,4-glucanases/cellulases (EGL, CEL) (del Campillo and Bennett, 1996), and expansins (EXP) (Belfield et al., 2005; reviewed by Kim, 2014). Recently, *BOP* expression (Frankowski et al., 2015) and the influence of abscisic acid (ABA) on the ethylene (ET) biosynthesis pathway (Wilmowicz et al., 2016) in the abscission zone of yellow lupine's generative organs have been described, but more detailed information about the molecular mechanisms underlying the plant organ abscission is still needed.

Next-generation sequencing (NGS), for example by using the Illumina platform, has become an efficient and powerful approach for functional genomic research, especially when working with non-model plants (Unamba et al., 2015). Recently, RNA-Seq has been widely used for global gene expression profile analyses of plant response to a variety of biotic and abiotic stresses, such as salt (Zhou et al., 2016) or cold (Wang et al., 2015). It has also been used to study the abscission of soybean leaves (Kim et al., 2015, 2016), tomato flowers (Sundaresan et al., 2016) and apple fruits (Ferrero et al., 2015).

In this study, we explored complex expression changes during flower and pod abscission in *de novo* assembled transcriptome datasets generated from six different libraries from three organ types of *L. luteus* cv. Taper. We aimed to isolate the unique transcripts and unigenes from flowers, flower pedicels and pods that undergo abscission and those that do not, and to identify the differentially-regulated genes (DEGs) involved in abscission. The main objective of our research was to improve our understanding of the molecular mechanism of generative organ abscission in yellow lupine and to find the common genes (a potential target for crop improvement) regulating this process.

MATERIALS AND METHODS

Details of the experimental design and RNA Seq data analysis pipeline are summarized in **Figure 1**.

Plant Material

For RNA-Seq analyses, different organs (flowers, pods and flower pedicels) from the yellow lupine (*L. luteus*) cultivar Taper cultivated in the growth chamber as previously described

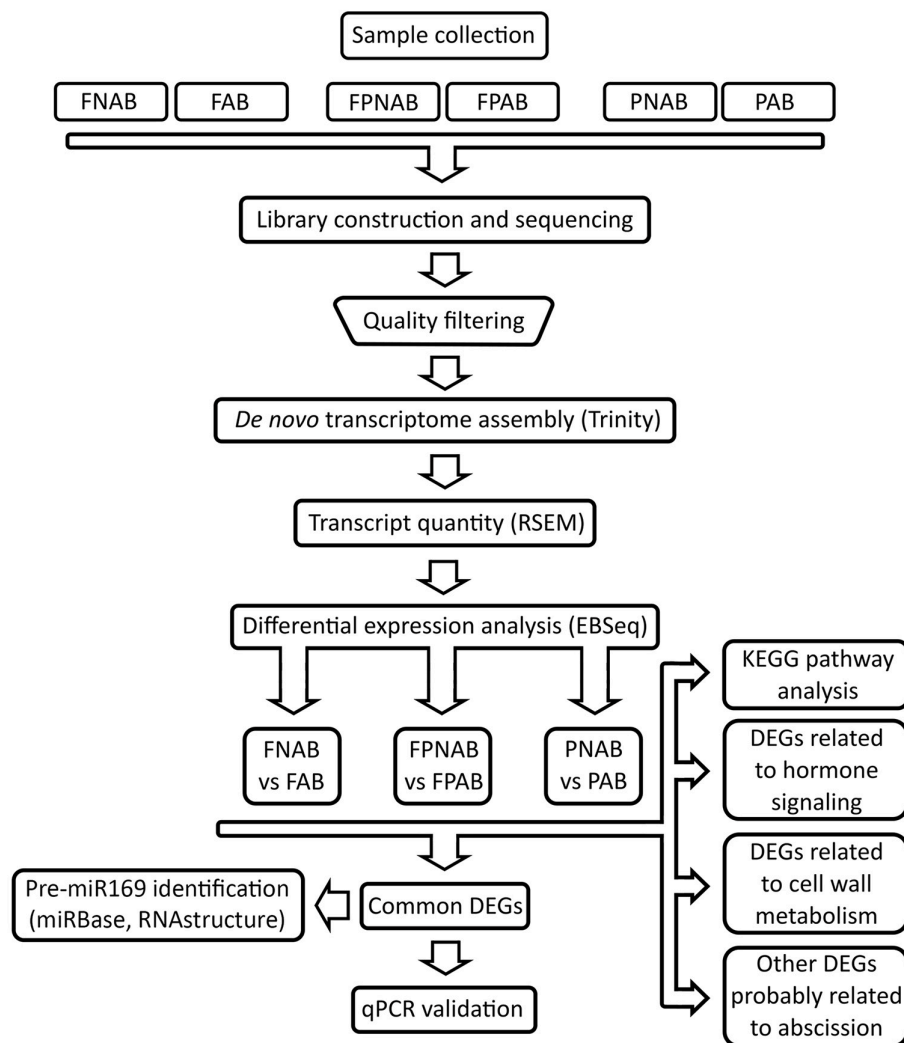


FIGURE 1 | Flowchart representing pipeline for RNA-seq experiments analysis conducted in the study. FAB vs. FNAB-abscising flowers and non-abscising flowers, FPAB vs. FPNAB-abscising flower pedicels and non-abscising flower pedicels, PAB vs. PNAB-abscising pods and non-abscising pods.

(Frankowski et al., 2015) were collected (**Figure 2**). From 54-day-old plants, whole fully opened flowers in stage 7 (Frankowski et al., 2015) with no signs of abscission (non-abscising flowers FNAB) were harvested (**Figure 2D**). Subsequently, their entire pedicels with an inactive abscission zone AZ located at the base of the flower pedicel (**Figures 2D,F**) were detached and collected as a separate sample (non-abscising flower pedicels FPNAB). From plants of the same age (54 days old), whole fully opened flowers with dehydrated petals (abscising flowers FAB) were harvested (**Figure 2E**). As above, their yellow and dehydrated pedicels with an active AZ (**Figures 2E,G**) were detached and collected as a separate sample (abscising flower pedicels FPAB). From 75 days-old plants pods with pedicels containing an inactive AZ (non-abscising pods PNAB) were harvested separately from the pods with an active AZ (abscising pods PAB) (**Figures 2H–K**). Following the harvest, the plant material was immediately frozen in liquid nitrogen and

stored at -80°C until the RNA isolation procedures could be started.

To confirm localization and activation of AZ in harvested samples, we cut longitudinally five randomly taken flower and pod pedicels from each variant and stained them with phloroglucinol-HCl (**Figures 2F,G,J,K**) according to Tadeo and Primo-Millo (1990). We applied a saturated solution of phloroglucinol (Sigma-Aldrich) in 20% HCl directly to samples. Then images were taken using CX31 microscope (Olympus, Japan), SC50 camera and Olympus cellSens Entry 1.14 software (Olympus, Japan).

cDNA Library Construction and Illumina Sequencing

Total RNA was isolated using the ISOLATE II RNA Plant Kit (Bioline, UK), following the manufacturer's protocol. The



FIGURE 2 | Generative organs of *Lupinus luteus* cv. Taper. Inflorescence development (A–C). Developing and abscising flowers, (D,E), respectively. Localization of harvested flower pedicels and abscission zones (AZ) indicated by squares and arrows, respectively. Cross-section view of developing and abscising flower pedicels after Phloroglucinol–HCl staining, (F,G), respectively. View of developing and abscising pods, (H,I), respectively. Cross-section view of developing and abscising pod pedicels after Phloroglucinol–HCl staining, (J,K), respectively. Black line – 1 cm bar for (A–E, H,I), 600 nm bar for (F,G,J,K).

concentration and purity were measured with the ND-1000 spectrophotometer (NanoDrop, USA) before proceeding with further analyses. The integrity and quality of total RNA was evaluated using the Bioanalyzer 2100 (Agilent 2100 Technologies, USA) and agarose gel electrophoresis. Two independent isolations were performed from at least 5 plants (25 flowers, pedicels or pods) for each variant. For further procedures we used RNA samples with RNA integrity number (RIN) >8: two for both variants of pedicels (2 FPAB and 2 FPNAB) and one each for abscisive and nonabscisive flowers and pods (1 FAB, 1 FNAB, 1 PAB, 1 PNAB). Both the cDNA library construction, and the transcriptome sequencing performed on the HiSeq 2000 platform (Illumina Inc., USA), were carried out by Genomed (Poland). The sequence data generated in this study has been deposited at NCBI, the Sequence Read Archive (SRA) database under the accession number PRJNA285604 (BioProject), with the experiment accession number SRX1069734.

Raw Sequence Processing and *De novo* Assembly

The strategy used for data processing and differential gene expression analysis is presented in Figure 1. The preliminary sequence analysis, including adapter removal and low quality overhang deletion was, performed using Trimmomatic software (Fragkostefanakis et al., 2012). Trimmomatic settings were optimized for paired-end reads. After adapter removal and

filtering out low quality reads, sequences shorter than 30 nt were discarded. The remaining sequences were analyzed using Trinity software (Grabherr et al., 2011) in accordance with the RNA-Seq analysis protocol published in Nature Protocols (Haas et al., 2013). At first, assembly based on *Lupinus angustifolius* (available at NCBI ID: 11024) (Yang et al., 2013) was considered, but due to the poor quality of its genome (191 thousands contigs) the idea was abandoned. *De novo* transcriptome assembly was performed using Trinity (Grabherr et al., 2011). The reference transcript assembly was obtained by joining individual reads into consensus sequences, which was then followed by the grouping of similar contigs into clusters using de Bruijn graphs. Based on these graphs, genes and their isoforms (single or multiple) were obtained. Among the 219,514 reference sequences obtained (with a mean length of 418 bp), 166,477 genes were discovered. Each isoform has been given an id number consisting of the cluster number (c), the number of genes within the given cluster (g) and the isoform number (i). For example, c10_g3_i1 is the first isoform of the third gene in the tenth cluster. The final results were indexed in the R environment (R Development Core Team, 2008).

Identification of Differentially Expressed Genes

Reads from each separate sample were mapped as a collection of reference transcripts using the Bowtie aligner, and the quantity of

each transcript in the sample was estimated using RSEM (part of the Trinity package) (Haas et al., 2013), using default settings for paired reads. The expression level was estimated at both unigene and isoform levels. Expression was described in three different ways: (1) the expected count: the number of reads per unigene, (2) TPM: the number of reads per unigene normalized to the library size (*Transcripts Per Million*), and (3) FPKM: the number of reads per unigene normalized to the library size and transcript length (*Fragments Per Kilobase Of Exon Per Million Fragments Mapped*).

Gene expression comparisons were performed using EBSeq software from the Bioconductor package (Leng and Kendzierski, 2015). This analysis was performed at the unigene level using the “expected count” expression type data. The following comparisons were made: FAB vs. FNAB, FPAB vs. FPNAB and PAB vs. PNAB. The obtained files were then filtered, leaving only those records where PPDE (Posterior Probability of Differential Expression) was less than 0.05 and $\text{Log}_2\text{FC} > 2$.

The sequence of each differentially expressed gene (DEG) was translated to the protein sequence, and then UniProt database (UniProt Consortium, 2015) was searched for sequences of similar proteins, and the best-matching protein was selected. If the protein ortholog could not be found, the unigene was assigned a “no hit” tag or left empty.

Analysis of Enriched KEGG Pathways for DEGs

An analysis of overrepresented KEGG pathways (Kanehisa and Goto, 2000; Kanehisa et al., 2016) was performed using plantGSEA software, but due to the small number of statistically significant hits (data not shown) we decided to repeat the analysis using a different tool. Unfortunately, in the case of plants most of the popular tools cannot be used. We decided to perform the analysis manually, creating a series of Python scripts. The first step was to find proteins of a length of at least 30 amino-acids in the studied sequences, and for each gene only the longest protein was picked. The protein was not required to contain a start or stop codon, which was a reasonable assumption considering the fact that the investigated transcripts were much shorter (774 nt on average, compared to 1,557 nt in *A. thaliana* transcripts). Afterwards, we used the KAAS server (http://www.genome.jp/kaas-bin/kaas_main), which allowed us to assign orthologous groups from the KEGG database to the proteins and peptides that we had found. We used the following parameters: (i) the search method: BLAST, (ii) the species: *A. thaliana*, *G. max*, *V. vinifera*, *S. lycopersicum*, *O. sativa*, (iii) the method for finding potential homologs: SBH (single-directional best hit). Afterwards, we downloaded KEGG pathways for *Glycine max* (gmx) based on the KAAS server and on the obtained results, and performed an analysis of overrepresented KEGG pathways in the examined collection of genes (DEGs). For the purpose of this testing, the accurate Fisher test and FDR correction were used.

Pre-miR169 Identification

DEGs common for all of the compared transcriptomes that showed no homology to proteins were analyzed in respect

of non-coding RNAs. The Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) in the miRBase web database (<ftp://mirbase.org/pub/mirbase/12.0/>) was used to search for pre-miRNA and mature miRNA homologous sequences in these DEGs. The c125095_g1_i1 unigene has shown strong similarity to mature and stem-loop sequences of miR169 from various plant species. The unigene sequence was analyzed using Search miRBase Web software (Griffiths-Jones, 2004; Griffiths-Jones et al., 2006, 2008; Kozomara and Griffiths-Jones, 2011, 2014). RNAstructure software (<http://rna.urmc.rochester.edu/RNAstructureWeb>) was used to compute the secondary structure of the c125095_g1_i1 unigene sequence.

Quantitative Real-Time RT-PCR Analysis

A quantitative real-time RT-PCR (RT-qPCR) analysis was used to validate RNA-Seq results. For cDNA synthesis, 1 µg of total RNA was reverse-transcribed using the TranscriptMe Kit (DNA Gdansk, Poland). The synthesized cDNA samples were diluted 5 times prior to the qPCR. The qPCR was conducted using the KAPA Probe Fast Universal qPCR Kit (KAPA Biosystems, South Africa) following the manufacturer's protocol. Each reaction was prepared in a total volume of 20 µl consisting of 10 µl of 2 × KAPA Probe Fast qPCR Master Mix, 0.2 µM of each primer, 0.1 µM of a specific UPL probe and 5 µl of diluted cDNA. The real time qPCR was carried out on the LightCycler480 (Roche, Switzerland) under the following conditions: 95°C for 10 min, 45 cycles of 95°C for 10 s, 58°C for 30 s, and 72°C for 1 s. For each variant there were two biological and three technical replicates. *LLActin* was used as the reference gene. Primers and UPL (Universal Probe Library) probes used in the qPCR were designed using ProbeFinder version 2.48 (Roche, Switzerland) and are listed in **Table S1**. The data were analyzed using LightCycler480 software (Roche, Switzerland).

RESULTS

RNA-Seq and *De novo* Transcriptome Assembly

In order to obtain an insight into the process of generative organ abscission in yellow lupine we constructed transcriptome libraries: for flowers (FAB), flower pedicels (FPAB) and pods (PAB) undergoing abscission (with an active AZ), and the same for organs that did not fall (FNAB, FPNAB, PNAB). In order to verify whether AZ is active or inactive in chosen stage of flower and pod development, we analyzed their lignification pattern in pedicels by phloroglucinol staining—just before accomplishment of abscission, lignin accumulates within AZ (Clements and Atkins, 2001). In pedicels of not falling flowers and pods we observed no staining within AZ (only vascular bundles were stained), while organs undergoing abscission showed positive color reaction (**Figures 2F,G,J,K**). Because of similarities in flower and pod AZ structure, we decided to perform RNA-Seq analysis only for flower pedicels.

The HiSeq 2000 platform (Illumina) generated from 56,784,288 to 46,619,042 raw reads in each library, which accounted for an average of 6.96 Gb raw data for each

sample. About 98% of the raw reads were of high quality and, after Trimmomatic was run, about 80% of them were ultimately mapped (Table S2). The data were deposited in the National Center for Biotechnology Information's (NCBI) Short Read Archive (SRA) database under the accession number PRJNA285604.

After *de novo* transcriptome assembly using Trinity software a set of 219,514 sequences of unique transcripts of a minimal length of 201 bp, including 166,473 different unigenes (the statistics of this process is summarized in Table 1), was obtained. Figure 3 shows the size distribution of the assembled transcripts, with the average size being 774 bp.

Identification of DEGs in AB vs. NAB Flowers, Flower Pedicels, and Pods

In order to investigate molecular changes occurring during organ abscission in yellow lupine we compared transcriptomes of FAB vs. FNAB, FPAB vs. FPNAB PAB vs. PNAB transcriptomes. Literature data indicate, that hormonal changes and reorganization of cell wall structure play the most important role in organ abscission (Ascough et al., 2005; Corbacho et al., 2013; Estornell et al., 2013; Kim et al., 2015; Sundaresan et al., 2016) therefore in an analysis of the identified differentially expressed genes (DEGs) we focused on cell wall modification enzymes and hormone metabolism. We also took into account other DEGs

that may be involved in organ fate. Finally, we identified DEGs that are common for all library comparisons.

A bioinformatics analysis with EBseq of the transcriptomes revealed that during generative organ abscission in yellow lupine 5989 unigenes were either up- or down-regulated (PPDE < 0.05, FC > 2). The number of DEGs varied between library comparisons: 1343 in flowers (FAB vs. FNAB), 4.5% of which (293) were strongly affected (\log_2 ratio over 5); 2933 in flower pedicels (FPAB vs. FPNAB), 9.5% of which (308) were strongly affected; 1491 in pods (PAB vs. PNAB), 11.6% of which (128) were strongly affected (Figure 4A and Table S3). In abscising flowers, more unigenes were up-regulated (1018) than down-regulated (325), while in the flower pedicels and in the pods the trend was opposite: more unigenes were down-regulated (1,712 and 1,126, respectively) than up-regulated (1,221 and 365, respectively) (Figure 4A). The number of unique DEGs was the highest in the flower pedicels (2,503), being two times lower in the pods (1,283) and the lowest in the flowers (923). Flowers and pedicels shared 307, pedicels and pods 95, and pods and flowers 85 DEGs. Only 28 DEGs were common for all library comparisons (Figure 4B).

DEGs Involved in Cell Wall Metabolism in Abscising and Maintained Generative Organs of Lupine

In abscising flowers, flower pedicels and pods, genes encoding in cell wall hydrolyzing enzymes were mostly up-regulated.

In abscising flowers, 27 DEGs were involved in cell wall metabolism. These unigenes showed similarity to *EXPANSINS EXP* (3 DEGs), *XYLOGLUCAN ENDOTRANSGLUCOSYLASE / HYDROLASES XTH* (1), *ENDOGLUCANASES / CELLULASES EGL/CELL* (2), *POLYGALACTURONASES PG* (5), *BETA-GALACTOSIDASES BGal* (4), *PECTINESTERASES PME* (10), *PECTIN ACETYLESTERASES PAE* (1), and *PECTATE LYASES PEC* (1). Most of them (22 DEGs) were up-regulated, and only 5 down-regulated in FNAB (Table 2 and Table S4).

TABLE 1 | Statistics of the sequencing and assembly.

Items	Number
Number of transcripts	219,514
Number of unigenes	166,473
Assembly length (Mb)	170
Minimum transcript length (bp)	201
Average transcript length (bp)	774
N50 (bp)	1,350

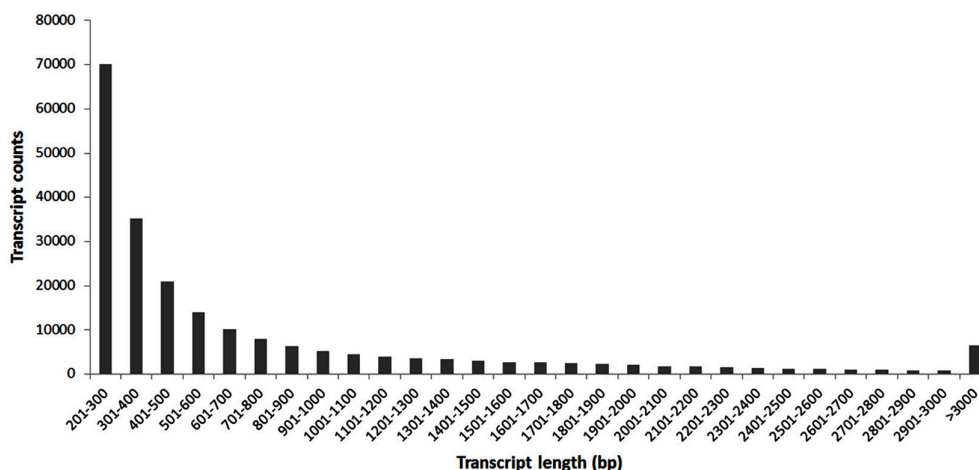


FIGURE 3 | The length distribution of the assembled transcripts.

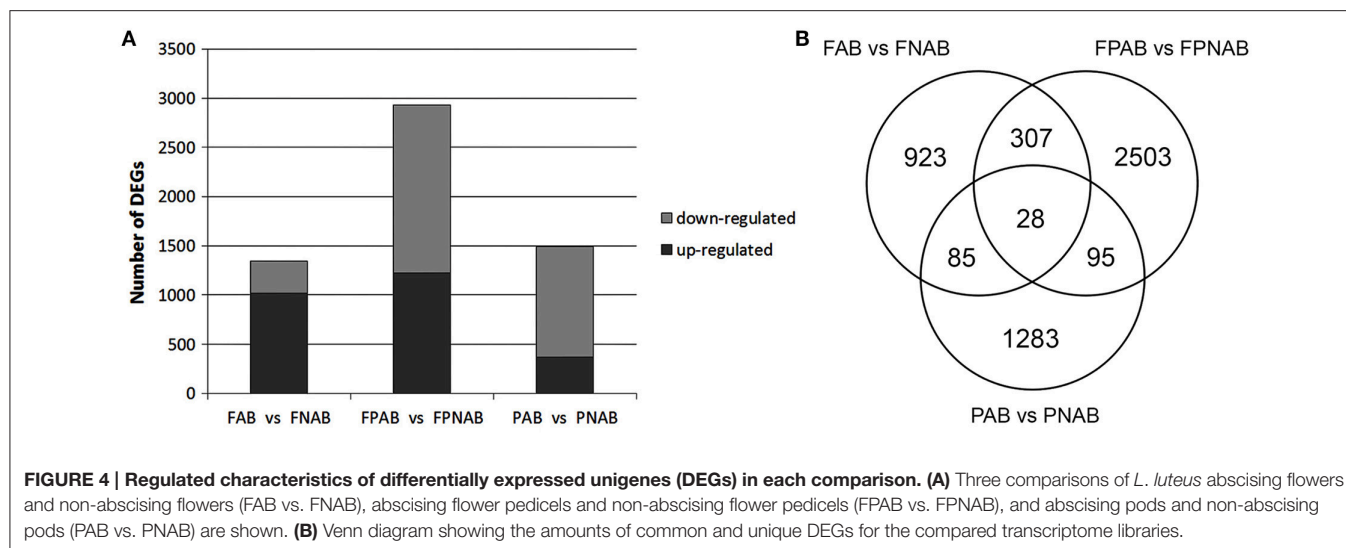


TABLE 2 | Differential expression patterns of cell wall-related genes in comparison of libraries FAB vs. FNAB, FPAB vs. FPNAB, and PAB vs. PNAB.

Gene description	FAB vs. FNAB		FPAB vs. FPNAB		PAB vs. PNAB	
	Up	Down	Up	Down	Up	Down
Expansin EXP	3	–	9	2	1	4
Xyloglucan endotransglucosylase/hydrolase XTH	1	–	9	9	1	4
Endoglucanase/Cellulase EGL/CELL	1	1	4	3	2	1
Polygalacturonase PG	5	–	18	7	–	10
Beta-galactosidase BGal	3	1	12	8	–	2
Pectinesterase PME	7	3	19	1	3	38
Pectin acetyltransferase PAE	1	–	2	2	–	2
Pectate lyase PEC	1	–	8	–	–	8

Numerous DEGs (113) were involved in cell wall metabolism in the flower pedicels. The identified unigenes showed similarity to *EXP* (11 DEGs), *XTH* (18), *EGL/CELL* (7), *PG* (25), *BGal* (20), *PME* (20), *PAE* (4), and *PEC* (8). Most of them (81) were up-regulated, and only 32 down-regulated in FPNAB (Table 2 and Table S4).

In abscising pods, 76 DEGs were involved in cell wall metabolism. The identified unigenes showed similarity to *EXP* (5 DEGs), *XTH* (5 DEGs), *EGL* (3 DEGs), *PG* (10 DEGs), *BGal* (2 DEGs), *PME* (41 DEGs), *PAE* (2 DEGs), and *PEC* (8 DEGs). Most of them (69 DEGs) were down-regulated, and only 7 up-regulated in PNAB (Table 2 and Table S4).

Differential Regulation of Genes Involved in Hormone Signaling in Compared Generative Organs

We observed differences in expression of genes involved in distribution and sensitivity to phytohormones during abscission of generative organs.

Auxin

Plants adopt different strategies to change the distribution pattern of active form of auxin across the tissues. Our study on yellow lupine revealed, that in generative organs undergoing abscission, following molecular mechanisms take place: (i) decrease in expression of auxin biosynthesis enzymes (like in FAB), (ii) and/or decrease in expression of its catabolism enzymes (like in FAB, FPAB), (iii) decrease in expression of enzymes that release auxin from its conjugates (like in PAB), (iv) increase (like in FPAB) and/or decrease in (like in FAB, FPAB) expression of its transport proteins. This resulted in complex changes in expression of genes involved in auxin signal transduction in all studied tissues.

In abscising flowers, 11 DEGs were involved in auxin (IAA) signaling (Table 3 and Table S5). Two enzymes up-regulated in FNAB ensure the right level of the active form of this phytohormone: YUCCA8 (c61021_g1), a key enzyme in IAA biosynthesis (Won et al., 2011), and 2-oxoglutarate-dependent dioxygenase (DAO) (c134360_g1), which is essential for auxin catabolism and the maintenance of auxin homeostasis in reproductive organs (Zhao et al., 2013). Only one unigene in this library comparison was responsible for auxin transport, *TORNADO 2* (c39845_g1) (Cnops, 2006), and it was up-regulated in FNAB. There were 8 DEGs involved in auxin signal transduction, three of which were down-regulated. Two DEGs displayed similarity to auxin response factor *ARF4* (c62886_g4) and *ARF7* (c61288_g4), and were up- and down-regulated, respectively.

Among DEGs identified in all library comparisons, most of the unigenes associated with auxin signaling (44) were differently expressed in the pedicels of flowers (Table 4 and Table S6). In the datasets for pedicels of the falling and the control flowers we found two DEGs responsible for auxin break-down: DAO (c134360_g1), essential for IAA oxidation, and indole-3-acetate O-methyltransferase 1 (c33445_g1), involved in methylation of the free carboxyl end of IAA. Both of them were over-expressed in FPNAB. Three of the four DEGs encoding transport proteins, namely auxin transporter-like protein *LAX4* (c61059_g4) and *LAX5* (c61634_g1, c11165_g1), were up-regulated in FPNAB.

TABLE 3 | Differential expression patterns of plant hormone metabolism and signaling-related genes in comparison of libraries FAB vs. FNAB.

Hormone	Total no. of DEGs	Up-regulated	Down-regulated	Function
IAA	11	1	–	Biosynthesis
		1	–	Catabolism
		1	–	Transport
		5	3	Signal transduction
GA	8	3	–	Biosynthesis
		–	2	Catabolism
		3	–	Signal transduction
ET	3	2	1	Signal transduction
ABA	2	–	2	Signal transduction
JA	3	2	–	Biosynthesis
		–	1	Signal transduction
CK	1	1	–	Catabolism
BR	2	2	–	Signal transduction
SA	1	1	–	Signal transduction
SL	1	1	–	Signal transduction

TABLE 4 | Differential expression patterns of plant hormone metabolism and signaling-related genes in comparison of libraries FPAB vs. FPNAB.

Hormone	Total no. of DEGs	Up-regulated	Down-regulated	Function
IAA	44	2	–	Catabolism
		4	1	Transport
		22	15	Signal transduction
GA	12	3	1	Biosynthesis
		–	4	Catabolism
		4	–	Signal transduction
ET	17	–	3	Biosynthesis
		8	6	Signal transduction
ABA	9	1	–	Catabolism
		–	8	Signal transduction
JA	3	–	3	Biosynthesis
CK	9	2	–	Biosynthesis
		3	3	Catabolism
		1	–	Signal transduction
BR	8	3	2	Biosynthesis
		3	–	Signal transduction
SA	6	2	4	Signal transduction

In this dataset we found two small auxin upregulated RNAs (SAURs), namely an up-regulated AGR7 (c50003_g2) and a down-regulated AGR2 (c59673_g5) in FPNAB. The list of DEGs identified in the pedicels was rich in genes encoding elements of auxin signal transduction. Two of the DEGs encoded receptors, namely a down-regulated F-box protein SKP2A (c54931_g4) and an up-regulated auxin-binding protein ABP19a (c58059_g1, c110842_g1, c52095_g1), three other DEGs encoded down-regulated ARF4 (c62886_g2), ARF7 (c128482_g1, c30002_g1) and ARF19 (c60211_g6), while as many as 15 DEGs encoded up- or down-regulated Auxin/Indole-3-Acetic Acid (AUX/IAA) proteins.

Among the genes differentially expressed in the abscising and non-abscising pods (**Table 5** and **Table S7**) we found one that encoded IAA-amino acid hydrolase ILL2 (c28519_g1) that catalyzes the hydrolysis of IAA conjugates and releases the active form of this hormone (LeClere et al., 2002) and was up-regulated in PNAB. Genes encoding auxin receptors TRANSPORT INHIBITOR RESPONSE 1 (TIR1) (c85345_g1), F-box protein SKP2A (c54931_g4) and auxin-binding protein ABP19a (c52095_g1, c58059_g1) were down-regulated. Simultaneously, two SAURs were down-regulated and one AUX/IAA was up-regulated in this dataset. Interestingly, two ARFs (5 DEGs), i.e., ARF2 (c52420_g3, c52420_g4, c52420_g1) and ARF4 (c62886_g7, c62886_g2), were significantly up-regulated in PNAB.

TABLE 5 | Differential expression patterns of plant hormone metabolism and signaling-related genes in comparison of libraries PAB vs. PNAB.

Hormone	Total no. of DEGs	Up-regulated	Down-regulated	Function
IAA	14	1	–	Catabolism
		6	7	Signal transduction
GA	12	9	–	Biosynthesis
		1	–	Catabolism
		–	2	Signal transduction
ET	5	–	3	Biosynthesis
		–	2	Signal transduction
ABA	2	1	–	Catabolism
		1	–	Signal transduction
JA	6	4	–	Biosynthesis
		1	1	Signal transduction
SA	6	1	5	Signal transduction
SL	1	1	–	Signal transduction

Ethylene

We observed differences in expression of ethylene (ET) biosynthesis genes in flower pedicels and pods and of genes involved in ethylene signal transduction in all studied organs.

In flowers we found 3 DEGs involved in ET signal transduction, among them up-regulated Mitogen-activated protein kinase 6 (c65488_g1), earlier proven to be involved in leaf senescence (Zhou et al., 2009) (**Table 3** and **Table S5**).

In the flower pedicels, 17 of the identified DEGs were involved in ET metabolism. Three unigenes encoding ET biosynthesis enzyme 1-aminocyclopropane-1-carboxylate oxidase (ACO) 1 (c60801_g4, c60409_g3, c140425_g1) were down-regulated in FPNAB. Eight of the fourteen genes encoding elements of ET signal transduction were up-regulated (**Table 4** and **Table S6**).

In pods all DEGs attributed to ethylene biosynthesis encoding 1-aminocyclopropane-1-carboxylate synthase 1 (ACS1) (c60313_g3, c58986_g1) and ACS7 (c107301_g1) and to signal transduction encoding ethylene-responsive transcription factor 3 (ERF3) (c18833_g1) and ERF1A (c81001_g1) were less intensively expressed in PNAB (**Table 5** and **Table S7**).

Gibberellin

We observed differences in expression of genes involved in gibberellin (GA) (i) biosynthesis: they were down-regulated in FAB and PAB, and mostly down-regulated in PFAB, except for the enzyme that catalyzes the transformation of direct precursor to GA active form (ii) catabolism: they were up-regulated in FAB and FPAB, and down-regulated in PAB (iii) signal transduction: they were down-regulated in FAB and up-regulated in PAB, and (iv) genes controlled by this hormone (down-regulated in FPAB).

As regards gibberellin eight DEGs in flowers were involved in GA signaling (**Table 3** and **Table S5**). Two of them encode gibberellin 2-beta-dioxygenase 1 (GA2OX1) (c57884_g1, c25173_g1), which is engaged in GA catabolism and were down-regulated in FNAB. The rest of them were up-regulated and included genes encoding enzymes responsible for GA biosynthesis: ent-kaurenoic acid oxidase 2 (KAO2) (c78744_g1), gibberellin 20 oxidase 2 (GA20OX2) (c54588_g2), GA20OX1 (c54588_g1); and for signal transduction: GAI (c49118_g1), gibberellin-regulated protein 12 (c59846_g1), and 14 (c48197_g2).

In flower pedices we found four DEGs engaged in GA biosynthesis (**Table 4** and **Table S6**). Three of them, namely GA20OX2 (c54588_g2), GA20OX1 (c54588_g1), and KAO2 (c60896_g9), were up-regulated, while the remaining one, LE (c24709_g1), was down-regulated in FPNAB. Three DEGs encoding the GA2OX1 enzyme (c25173_g1, c57884_g1, c49763_g3) responsible for hydroxylation of the active form of GA (Martin et al., 1999), were down-regulated. Also, four DEGs similar to gibberellin-regulated proteins were up-regulated in FPNAB.

All 9 DEGs attributed to GA biosynthesis were much more intensively expressed in PNAB (**Table 5** and **Table S7**). For example, the c59646_g2 unigene, GA20OX2 with $\log_2FC = 12.88$, was the most up-regulated DEG in PNAB. Only one DEG involved in GA catabolism encoding GA2OX1 (c49763_g2) was up-regulated on a moderate level. DEGs involved in GA response encoding GA signal transduction transcription factor GAMYB (c40131_g1, c60985_g11) were significantly down-regulated in PNAB.

Abscisic Acid

We observed differences in expression of genes involved in abscisic acid (ABA) catabolism (they were down-regulated in FPAB and PAB), and signal transduction (they were up-regulated in FAB and FPAB).

Two DEGs down-regulated in FNAB were involved in ABA signaling (**Table 3** and **Table S5**). That DEGs indicate similarity to PYL2 (c35977_g2 and c35977_g3) that are responsible for ABA perception (Yin et al., 2009).

Among the DEGs found in the flower pedicels and recognized as involved in ABA signaling (**Table 4** and **Table S6**), PYL1 (2 DEGs), SAPK2 (1 DEG), SAPK8 (1 DEG) and PP2CA (4 DEGs) were down-regulated in FPNAB. Simultaneously, abscisic acid 8'-hydroxylase 4 CYP707A4 (c61127_g1) engaged in ABA breakdown was up-regulated.

Additionally, ABA catabolism-related gene encoding abscisic acid 8'-hydroxylase 1 (CYP707A1) (c40794_g1) that catalyzes the first step of ABA degradation was up-regulated in PNAB (**Table 5** and **Table S7**).

Jasmonate

We observed differences in expression of genes involved in jasmonate (JA) biosynthesis (they were down-regulated in FAB) and signal transduction (they were up-regulated in FAB, FPAB and down-regulated in PAB).

We found three DEGs associated with JA metabolism in flowers. Two of them were up-regulated in FNAB, and these were chloroplastic LIPOXYGENASE 6 (LOX6) (c92415_g1) and JASMONATE O-METHYLTRANSFERASE (c121732_g1) that were involved in jasmonic acid (JA) and ester methyl jasmonate (JaMe) biosynthesis, respectively. The other one was more frequently expressed in FAB, namely TIFY10A (c52719_g1) engaged in JA signal transduction (**Table 3** and **Table S5**).

Three DEGs encoding chloroplastic JA biosynthesis enzymes, namely LINOLEATE 9S-LIPOXYGENASE 5, LOX2 (c45959_g2, c57820_g1) and allene oxide synthase CYP74A (c92570_g1) were down-regulated in FPNAB (**Table 4** and **Table S6**).

In contrast, DEGs associated with JA biosynthesis were all up-regulated in PNAB. The DEGs associated with JA biosynthesis, i.e., JASMONATE O-METHYLTRANSFERASE (c43486_g2), LOX1 (c92206_g1, c70262_g1) and LOX4 (c61214_g3), were up-regulated in PNAB (**Table 5** and **Table S7**).

Brassinosteroid

Brassinosteroid (BR) signaling decreased in abscising flowers and their pedicels. In flower pedicels there were complex differences in expression of BR biosynthesis enzymes.

Two genes involved in BR signal transduction and one engaged in response to this hormone were significantly down-regulated in flowers (**Table 3** and **Table S5**).

Among the 8 DEGs in the flower pedicels recognized as involved in brassinosteroid (BR) signaling, all the three involved in signal transduction were up-regulated in FPNAB (**Table 4** and **Table S6**). The expression profile of DEGs putatively encoding BR biosynthesis enzymes was more complex: some of them were up- and others were down-regulated. Unigenes encoding the same protein, cytochrome P450 85A

(CYP85A), and belonging to the same cluster were differentially expressed: c54957_g3 and c54957_g2 were up-regulated, while c54957_g1 was down-regulated in FPNAB. Other DEGs putatively encoding BR biosynthesis enzymes, namely 3-epi-6-deoxocathasterone 23-monooxygenase CYP90D1 (c51134_g1) and CYP85A1 (c57767_g3), were up- and down-regulated in FPNAB, respectively.

In pods there was not find any unigene related to BR metabolism.

Cytokinin

We observed differences in expression of genes involved in cytokinin (CK) (i) biosynthesis: they were down-regulated in FPAB, (ii) catabolism: they were down-regulated in FAB and either down- and up-regulated in FPAB, and (iii) signal transduction: they were down-regulated in FPAB.

In FAB vs. FNAB transcriptome comparison we found one DEG related to CK metabolism the gene encoding cytokinin hydroxylase CYP735A1 (c59901_g1) involved in CK catabolism (Table 3 and Table S5).

In flower pedicels (Table 4 and Table S6), among the 9 DEGs related to cytokinin (CK), two genes encoding enzymes that catalyze the final steps of CK biosynthesis, cytokinin riboside 5'-monophosphate phosphoribohydrolases LOG1 (c62014_g7) and LOG3 (c44862_g1) (Kuroha et al., 2009), were significantly up-regulated in FPNAB. Also, genes encoding cytokinin dehydrogenases that catalyze CK oxidation to a biologically inactive form displayed differential expression: two of them, CKX5 (c62634_g1) and CKX3 (c55747_g3, c55747_g1), were up-regulated, while the other two, namely CKX1 (c50912_g1, c50912_g4) and CKX9 (c108053_g1), were down-regulated in FPNAB. We found only one DEG assigned to cytokinin signal transduction representing the CK receptor, namely histidine kinase *AHK4* (c49543_g4), and it was up-regulated in FPNAB.

In pods there was not find any unigene related to CK metabolism.

Salicylic Acid

Expression of genes involved in salicylic acid (SA) signal transduction was down-regulated in abscising flowers, and in other studied organs these differences were more complex.

One gene participating in salicylic acid signal transduction, namely pathogenesis-related protein PR-1 (c16844_g1), was up-regulated in FNAB (Table 3 and Table S5).

Among the DEGs identified in the flower pedicels as involved in salicylic acid signaling we found five unigenes (encoding 4 proteins) showing similarity to elements of signal transduction: one up-regulated pathogenesis-related protein PR1 (c16844_g1) and three down-regulated ones in FPNAB: transcription factor TGA1 (c8716_g1 and c53737_g2), Salicylic Acid-Binding Protein 2 (c93153_g1), and PRB1 (c33973_g2) (Table 4 and Table S6).

One of the six DEGs attributed to SA-related genes and belonging to PR1 gene family were over-expressed in PNAB, while the remaining five were down-regulated (Table 5 and Table S7).

Stigolactone

In flowers and pods transcriptome comparisons we found one DEG associated with strigolactone signal transduction (up-regulated in FNAB and PNAB), strigolactone esterase DAD2 (c47746_g1) that binds and hydrolyzes mobile strigolactones initiating SCF-mediated strigolactone signal transduction (Hamiaux et al., 2012) (Tables 3, 5, Tables S5, S7).

Characterization of Other DEGs Probably Related to Generative Organ Abscission

Characterization of DEGs from Abscised Flowers and Control

Among DEGs in flowers we also found genes encoding proteins directly involved in protection against reactive oxygen species (ROS) (Table S8), namely peroxidase 12 (c95461_g1, log₂FC = 4.2) and peroxidase 10 (c108985_g1, log₂FC = -4.6), as well as enzymes belonging to the carotenoid biosynthesis pathway, namely chloroplastic carotenoid cleavage dioxygenase 4 (c63538_g3, log₂FC = 4.9), chloroplastic or chromoplastic zeta-carotene desaturase (c23470_g1, log₂FC = 4.5) and chloroplastic beta-carotene hydroxylase 2 (c47712_g2, log₂FC = 3.9).

The flowers that are dropped and those that are maintained on the plant also differ in respect of the genes involved in proper fertilization (Table S8). Some DEGs, such as NUCLEAR FUSION DEFECTIVE 4 (NDF4) (c41239_g1) encoding a protein required for fusion of polar nuclei during female gametophyte development and karyogamy during fertilization (Portereiko et al., 2006), or POLLENLESS3 (c47977_g1) which is essential for male fertility, especially for microspore and pollen grain production (Glover et al., 1998), were less abundantly expressed in FAB.

Characterization of Other DEGs from Flower Pedicels with Active AZ and Control

As it was in the case of flowers, we found DEGs encoding elements engaged in protection against ROS, and these were 18 different peroxidases (10 up- and 8 down-regulated) and chloroplastic beta-carotene 3-hydroxylase (c57645_g2, log₂FC = -2.7) (Table S9).

Additionally, one of the most intensively expressed DEGs in the flower pedicels with an active AZ showed similarity to vacuolar processing enzyme VPE (8 unigenes, log₂FC from -2.69 to -7.09), an executor of plant PCD (Programmed Cell Death) (Hatsugai et al., 2004) (Table S9). VPE is a cysteine protease that cleaves a peptide bond at the C-terminal side of asparagine and aspartic acid (Wang et al., 2009) and plays an essential role in the regulation of the lytic system of plants during their defensive and developmental processes (Hara-Nishimura et al., 2005).

In the flower pedicels, we identified several transcription factors specific for soybean leaf and flower abscission (Kim et al., 2016), such as NAC (29 DEGs), WRKY (12 DEGs) and AIL/PLETHORA (PLT) (3 and 2 DEGs respectively) which were up-regulated in FPAB (Table S9).

Interestingly, transcripts of 12 DEGs encoding aquaporins belonging to tonoplast intrinsic protein (TIP) and plasma membrane intrinsic protein (PIP) gene families were more accumulated in the pedicels of non-abscising flowers (Table S9).

Characterization of Other DEGs from Abscising and Maintained Pods of Lupine

Similarly to the case of flowers and flower pedicels, we found DEGs encoding 8 different peroxidases, 3 up- and 5 down-regulated, but no enzyme belonging to the carotenoid biosynthesis pathway (Table S10).

Other DEGs with a known homology, apart from the one that was the most up-regulated in the pods (gibberellin 20 oxidase), were those that showed similarity to genes encoding arabinogalactan proteins, i.e., PISTIL-SPECIFIC EXTENSIN-LIKE protein (c33p206_g1, $\log_2FC = 9.5$) and FASCICLIN-LIKE ARABINOGALACTAN PROTEIN 10 (FLA) (c141625_g1, $\log_2FC = 8.8$) (Table S10). In tomato these types of protein are preferentially expressed in immature white fruits (Fragkostefanakis et al., 2012).

During data interpretation special attention was paid to the AGL62 Agamous-like MADS-box protein (c59718_g1, $\log_2FC = 8.6$) (Table S10). The AGL62 suppresses cellularization during endosperm development in *Arabidopsis* (Kang et al., 2008).

Another DEG that could play a role in pod abscission or development is transcription factor MYC4 (c96574_g1, $\log_2FC = 5.25$; Table S10). Recent studies showed that MYC4 together with MYC2 and MYC3 are involved in regulating seed production in a JA-dependent manner (Qi et al., 2015).

By analyzing the same dataset we discovered, that in the pods undergoing abscission (PAB) as many as 26 unigenes related to calcium transport, homeostasis and response were over-expressed (\log_2FC varied from -2.86 to -4.68). This set of genes encompassed putative calcium-transporting ATPase (1 DEG), calcium-dependent protein kinases (6 DEGs), probable calcium-binding proteins (12 DEGs), cation/calcium exchanger (1 DEG) and calmodulin-like proteins (6 DEGs) (Table S10).

Common DEGs Identified for Comparisons of Maintained and Abscised Flowers, Flower Pedicels and Pods

We identified 28 DEGs that are common for all library comparisons. Almost all of them were regulated in the same direction in the flowers, flower pedicels and pods: 25 DEGs were down-regulated and two up-regulated in all comparisons. The only exception was TIP2:1 (c84944_g1), which was over-expressed in FPAB and less expressed in FNAB and PNAB (Table 6 and Table S11).

Among the common DEGs, there were representatives of factors involved in cell wall reorganization (*ENDOGLUCANASE 11*, *EGL11*), water transport (*TIP2:1*), stress response (for exp. *ZAT12*, *HSF24*, *RCI2B*, *INPP5A*, *SPX*, *MYB4*, or *SIB2*) and calcium signaling (*ACA13*). One from common DEGs showing similarity to *QUIRKY* (*QKY*) encoding a predicted membrane-bound protein.

Among the common DEGs, seven unigenes showed no homology to the protein sequence. A further analysis of these sequences revealed that the c125095_g1 DEG showed similarity to pre-miR169e from *Glycine max* (gma-MIR169e) (Figure 5). This unigene was down-regulated in all cases (Table 6 and Table S11).

RT-qPCR Validation of RNA-Seq Analysis

In order to confirm the differential expression profiles of the DEGs identified by the RNA-Seq analysis we selected 11 candidates from the set of DEGs common for three library pairs. These were 2 up-regulated and 8 down-regulated DEGs. The RT-qPCR-derived expression patterns of these DEGs fitted well with these determined by the RNA-Seq analysis (Figure S1).

The Enriched Pathways (KEGG) Analysis of DEGs from Compared Generative Organ Libraries

In order to categorize the biological functions of the identified DEGs, we performed a KEGG pathway enrichment analysis (Kanehisa and Goto, 2000; Kanehisa et al., 2016).

KEGG Pathway Analysis of DEGs from FAB vs. FNAB Comparison

In the flowers, 416 (including 353 up- and 63 down-regulated) DEGs were mapped to 40 KEGG pathways (p -value < 0.05) (Table S12). Most of the pathways (32 KEGGs) were part of the major category of “metabolism,” while the others were classified as “genetic information processing” (6 KEGGs), “environmental information processing” (1 KEGG) or “cellular processes” (1 KEGG) (Figure S2A).

Pathways with most numerous DEGs were mapped to the “global/overview maps” category that included “metabolic pathways” [gmx01100] (98 DEGs), “biosynthesis of secondary metabolites” [gmx01110] (68 DEGs) and “biosynthesis of antibiotics” groups [gmx01130] (29 DEGs) (Figure 6 and Figure S2A).

Within the “metabolism” category, 10 general pathways were distinguished, among which the most enriched ones were classified as “carbohydrate metabolism” [gmx00040, gmx00500, gmx00520] (36 DEGs), “amino acid metabolism” [gmx00400, gmx00350, gmx00290, gmx00360, gmx00330, gmx00280, gmx00380] (33 DEGs) and “metabolism of terpenoids and polyketides” [gmx00900, gmx00902, gmx00904, gmx00909, gmx00903, gmx00908, gmx00906] (30 DEGs) (Figure 6 and Figure S2B).

KEGG Pathway Analysis of DEGs from FPAB vs. FPAB Comparison

In the flower pedicels, 715 (including 277 up- and 438 down-regulated) DEGs were mapped to 63 KEGG pathways (p -value < 0.05) (Table S13). Most of the pathways (52 KEGGs) were part of the “metabolism” network, and the other ones are classified as “environmental information processing” (3 KEGGs), “cellular processes” (2 KEGGs), “genetic information processing” (1 KEGG) and “organismal systems” (1 KEGG) (Figure S3A).

Pathways with most numerous DEGs were mapped to “metabolic pathways” [gmx01100] (154 DEGs) and “biosynthesis of secondary metabolites” [gmx01130] (113 DEGs) from the “global/overview maps” category, and “plant hormone signal transduction” [gmx04075] (40 DEGs) from the “environmental information processing” category (Figure 7 and Figure S3A).

TABLE 6 | List and annotation of common DEGs identified in all three comparison of libraries FAB vs. FNAB, FPAB vs. FPAB, PAB vs. PNAB.

No	Unigene ID	FAB vs. FNAB		FPAB vs. FPAB		PAB vs. PNAB		Gene description
		log ₂ FC	Regulated	log ₂ FC	Regulated	log ₂ FC	Regulated	
1	c107017_g1	-7.17	Down	-4.94	Down	-2.87	Down	tRNA (adenine(34)) deaminase, chloroplastic
2	c107270_g1	-3.82	Down	-3.05	Down	-5.54	Down	No result
3	c107857_g1	-5.65	Down	-2.94	Down	-3.65	Down	Monogalactosyldiacylglycerol synthase 2, chloroplastic
4	c10793_g1	-5.12	Down	-3.23	Down	-3.69	Down	EG45-like domain containing protein
5	c121523_g1	-6.03	Down	-4.83	Down	-4.72	Down	No result
6	c125095_g1	-3.89	Down	-8.00	Down	-3.30	Down	MIR169e
7	c15548_g1	-7.11	Down	-4.21	Down	-3.57	Down	Alpha-aminoadipic semialdehyde synthase
8	c16039_g1	-4.01	Down	-3.19	Down	-2.87	Down	Putative calcium-transporting ATPase 13, plasma membrane-type
9	c27045_g1	-4.12	Down	-2.65	Down	-2.95	Down	Coumaroyl-CoA:anthocyanidin 3-O-glucoside-6''-O-coumaroyltransferase 1
10	c29048_g1	-4.10	Down	-3.72	Down	-2.94	Down	Zinc finger protein ZAT12
11	c3490_g1	-4.81	Down	-5.33	Down	-3.03	Down	Type I inositol 1,4,5-trisphosphate 5-phosphatase 11
12	c3661_g1	-4.34	Down	-7.71	Down	-3.83	Down	Sigma factor binding protein 2, chloroplastic
13	c36657_g1	-4.25	Down	-4.09	Down	-3.09	Down	Heat shock factor protein HSF24
14	c51152_g1	-4.14	Down	-3.89	Down	-3.30	Down	Hydrophobic protein RCI2B
15	c51204_g1	-3.99	Down	-2.98	Down	-3.11	Down	Heat shock factor protein HSF24
16	c51606_g1	-4.40	Down	-4.40	Down	-2.85	Down	Protein QUIRKY
17	c52215_g2	5.36	Up	3.52	Up	3.65	Up	14 kDa proline-rich protein DC2.15
18	c53596_g1	-6.50	Down	-4.06	Down	-2.73	Down	SPX domain-containing protein 3
19	c55641_g3	4.26	Up	3.54	Up	4.16	Up	Endoglucanase 11
20	c56313_g2	-4.32	Down	-3.69	Down	-3.46	Down	No result
21	c57393_g3	-3.91	Down	-5.54	Down	-2.85	Down	Myb-related protein Myb4
22	c58087_g1	-5.11	Down	-4.55	Down	-4.24	Down	Ubiquitin fusion degradaton protein
23	c58367_g2	-4.57	Down	-6.39	Down	-2.98	Down	Beta-glucosidase 12
24	c65572_g1	-4.00	Down	-5.71	Down	-3.08	down	No result
25	c82312_g1	-4.64	Down	-4.02	Down	-3.43	Down	No result
26	c84944_g1	-4.22	Down	2.50	Up	-2.64	Down	Aquaporin TIP2-1
27	c86078_g1	-6.68	Down	-2.59	Down	-3.32	Down	Inorganic pyrophosphatase 2
28	c94313_g1	-5.36	Down	-4.79	Down	-2.73	Down	No result

Within the “metabolism” category, we distinguished 9 general pathways, among which the most enriched ones were classified as “amino acid metabolism” [gmx00360, gmx00350, gmx00270, gmx00290, gmx00340, gmx00280, gmx00260, gmx00330, gmx00250, gmx00310] (83 DEGs), “carbohydrate metabolism” [gmx00500, gmx00040, gmx00562, gmx00520] (67 DEGs) and “lipid metabolism” [gmx00062, gmx01040, gmx00073, gmx00564, gmx00565, gmx01212, gmx00592, gmx00591, gmx00061, gmx00590] (66 DEGs) (Figure 7, Figure S3B).

KEGG Pathway Analysis of DEGs from PAB vs. PNAB Comparison

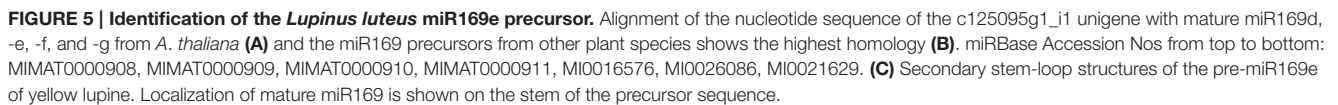
In the pods, 508 (including 114 up- and 394 down-regulated) DEGs were mapped to 30 KEGG pathways (p -value < 0.05) (Table S14). Most of the pathways (22 KEGGs) were part of the “metabolism” network, and the other ones were classified as “genetic information processing” (2 KEGGs), “environmental information processing” (1 KEGG) and “organismal system” (1 KEGG) (Figure S4A).

Pathways with most numerous DEGs were mapped to “metabolic pathways” [gmx01100] (133 DEGs) and “biosynthesis of secondary metabolites” [gmx01130] (67 DEGs) from the “global/overview maps” category, and “pentose and glucuronate interconversions” [gmx00040] (50 DEGs) from the “metabolism” category (Figure 8, Figure S4A).

Within the “metabolism” category, we distinguished 7 general pathways, among which the most enriched ones were classified as “carbohydrate metabolism” [gmx00040, gmx00500, gmx00053, gmx006562, gmx00052, gmx00010, gmx00520] (134 DEGs), “amino acid metabolism” [gmx00350, gmx00330, gmx00360, gmx00270] (40 DEGs) and “biosynthesis of other secondary metabolites” [gmx00950, gmx00970, gmx00960, gmx00940] (23 DEGs) (Figure 8, Figure S4B).

DISCUSSION

The process of flower abscission in yellow lupine was investigated already in late 1950s. Those studies demonstrated a significant influence of the flower’s location in the inflorescence (Van



Our RNA-Seq analyses of *L. luteus* generative organs showed bidirectional changes in the expression of the cell wall-related genes, probably associated not only with the ongoing process of abscission but also with progressive development of the organs that are not dropped (**Table 2**). In a flower transcriptome comparison, we identified 27 DEGs involved in cell wall function. Most of them were up-regulated in FNAB, and only 5 were down-regulated. This is probably related to the inhibition and continuation of growth in flowers being dropped and those remaining on the plant, respectively. This supposition is supported by two facts: (1) non-abscission lupine flowers at this stage of development showed the first signs of ovary growth (data not shown) and (2) pollination in tomato developing flower caused increase in the transcript levels of cell expansion-related genes like *SIEXPA5*, *SIPEC*, and *SIXTH1* (Vriezen et al., 2007; de Jong et al., 2011). Our results confirm our hypothesis that the abscising flowers are not fertilized, probably because they stopped developing properly even before pollination. In contrast, FNAB at this stage of development showed the first signs of ovary growth that required up-regulation of some cell wall modification-related genes. We

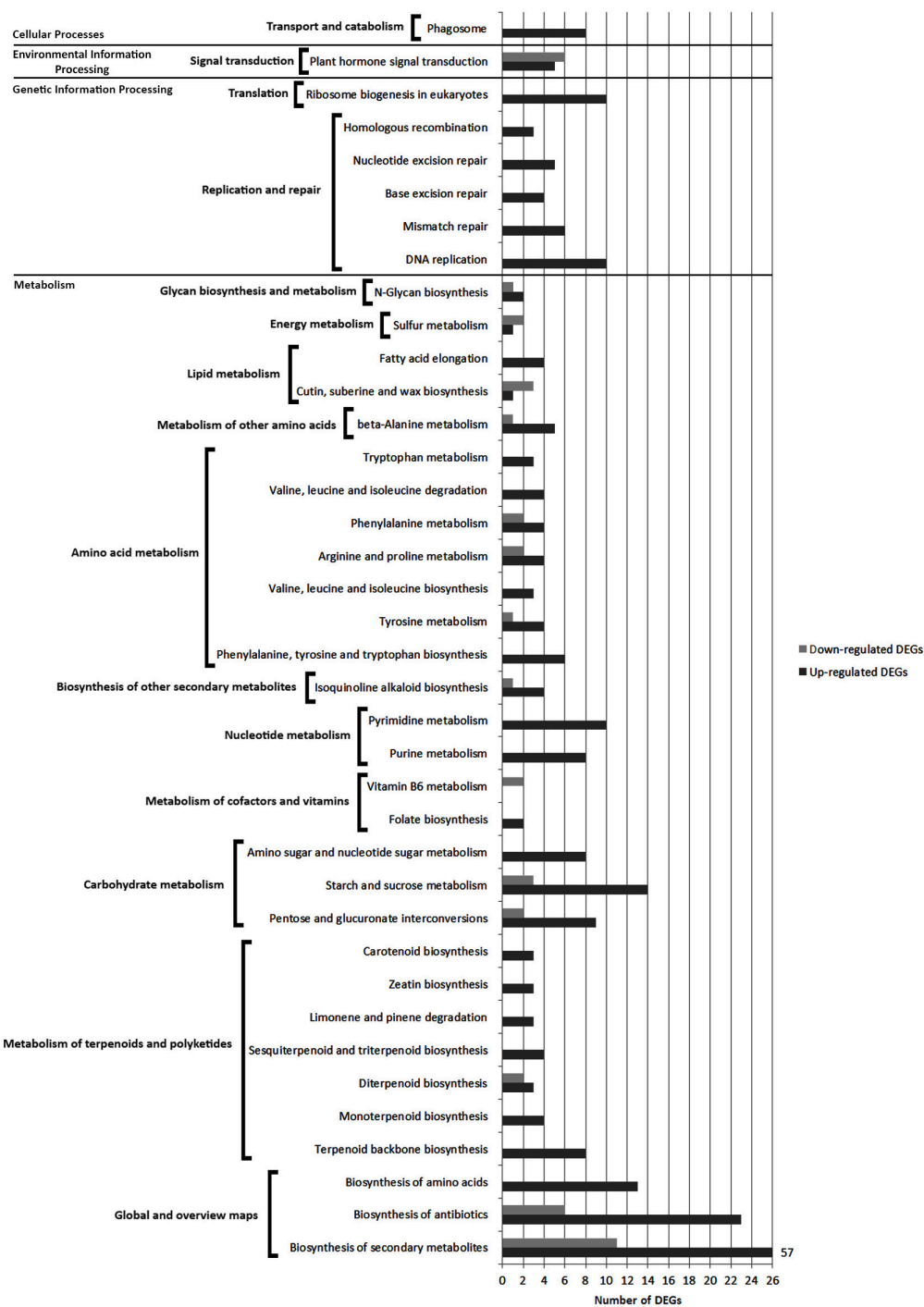


FIGURE 6 | Comparison of the number of up- and down-regulated DEGs identified in FAB vs. FNAB library comparison in the KEGG pathways.

strongly believe, that the lack of fertilization is caused by cues triggered from mature flowers on the first whorl and axillary flowers, that elicit modification of expression of specific genes during development of immature flowers. Verification of this hypothesis and identification of these cues need further study.

In our study, the samples were collected at the step of abscission activation (Kim, 2014) (visible lignin accumulation in AZ, **Figures 2G,K**), where biochemical changes in the cell walls within the AZ are crucial. Many studies (e.g., del Campillo et al., 1988; Lashbrook and Cai, 2008; Roongsattham et al., 2012; Estornell et al., 2013; Tsuchiya et al., 2015) indicate

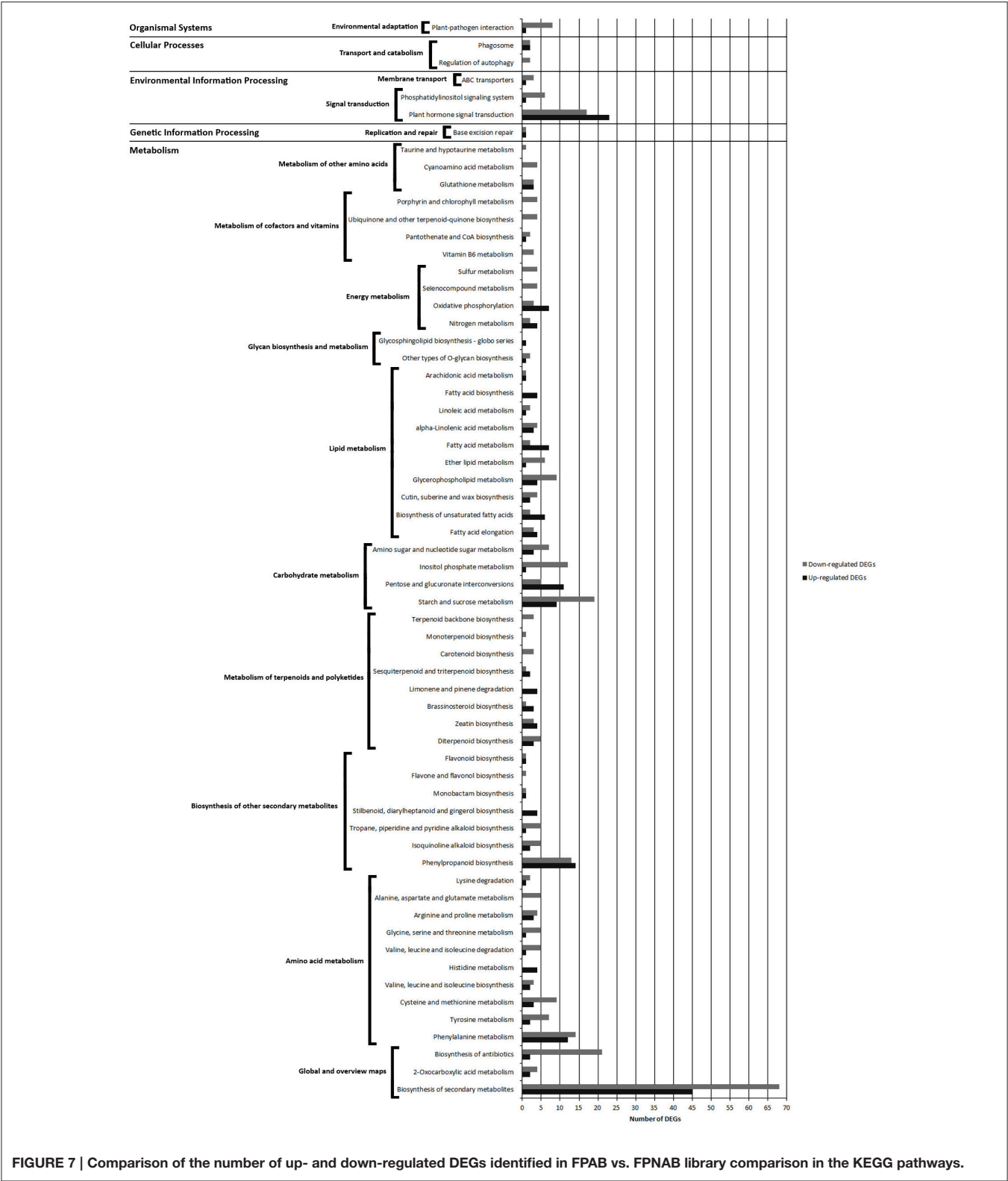


FIGURE 7 | Comparison of the number of up- and down-regulated DEGs identified in FPAB vs. FPNAB library comparison in the KEGG pathways.

the importance of *PG*, *CEL*, *XTH*, and *EXP* genes in organ abscission. Our RNA-Seq analyses showed that in flower pedicels many genes that encoded enzymes responsible for cell wall and middle lamella degradation and remodeling factors, which are the main targets at the late abscission stages, were differently expressed. These included *EXP*, *XTH*, *EDG/CEL*, *PG*, *BGal*, *PME*,

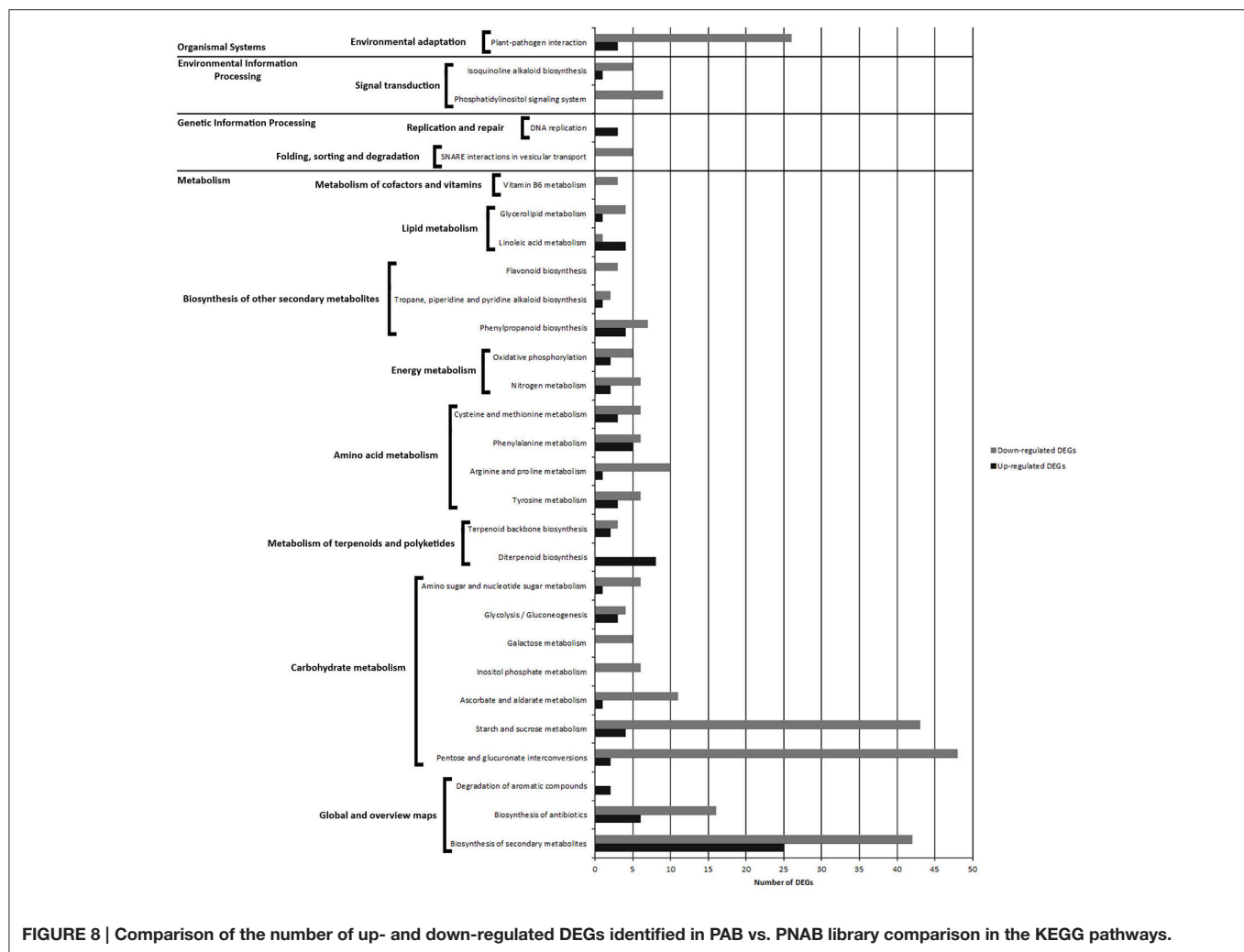


FIGURE 8 | Comparison of the number of up- and down-regulated DEGs identified in PAB vs. PNAB library comparison in the KEGG pathways.

PAE, PEC genes (Table 2 and Table S4), which were previously demonstrated to be highly expressed in the AZs of a large number of abscising organs (Lashbrook et al., 1994; del Campillo and Bennett, 1996; Cho and Cosgrove, 2000; Taylor and Whitelaw, 2001; Roberts et al., 2002; Ogawa et al., 2009; Meir et al., 2010). However, Tucker et al. (2007) discovered that in the soybean leaf AZ most of the cell wall-related genes displayed increased expression, while some of them showed no change or were down-regulated (e.g., *EXP1*, *PG6*, *PG7*, *PG16*, and *CEL11*). The authors suggested that different gene products may play very specialized roles in cell wall modifications and down-regulation of some of them may be necessary for the optimal functioning of the others (Tucker et al., 2007). In our case, the identified DEGs showed diverse expression patterns (they were either up- or down-regulated), which may be also related to the fact that we constructed the libraries by taking whole flower pedicels containing not only AZ cells, but also a large number of remaining stalk cells. Additionally, we compared pedicels of growing flowers. The identified DEGs may also be associated with continuation of flower development, which involves the strengthening of pedicels, a further development of vascular

bundles, etc. However, active AZ-specific genes (Estornell et al., 2013), namely *CEL1* (c33456_g1), *QRT2* (c56936_g4), *XTH2* (c48301_g1), and *BGAL1* (c59734_g1), were highly expressed in FPAB (Table S4) which indicate that in *L. luteus* organ abscission process are involved similar cell wall modification enzymes as in other plant species.

Pods are shed at the early stages of development (at a length of approximately 1 cm) (Figure 2) (Van Steveninck, 1958), and their location in the inflorescence does not inevitably determine their fate. Pods fall off from higher whorls, but some also from the lowest ones, where most of the pods remain on the plant. They probably compete in some way—the first pod to set has the best chance of achieving full development, while the last one to do so is dropped. Presumably, the signal from older pods causes developmental arrest of younger ones and induce organ abscission. A similar regulation mechanism of shedding fruits is suggested in apple (Botton et al., 2011). In pods of lupine studied here, 76 DEGs were attributed to cell wall metabolism and most of them were over-expressed in the abscising pods that indicate extensive degradation of cell wall structure in these organs (Table 2 and Table S4).

Hormonal Metabolism and Signaling—Related Genes Involved in Lupine Organ Abscission

Plant hormones are involved in regulating all developmental processes in plants (Arteca, 1996) and a large body of evidence indicates that hormones are the critical regulators of organ abscission (for a review see: Kim, 2014; Sawicki et al., 2015). Our analyses of DEGs showed that in yellow lupine genes involved in the auxin and gibberellins metabolism and action presented greatest expression changes, followed by changes in the expression of genes related to the functioning of ABA and ethylene. There were also differences in JA, CK, BR, SA, and STK metabolism.

Organs Abscission in Yellow Lupine is Associated with Intensive Changing of Auxin Catabolism and Signaling

Our RNA-Seq analyses indicate that the process of generative organ abscission in yellow lupine is associated with significant changes of auxin balance at the level of catabolism, perception, transport, and response to the phytohormone. The expression of genes encoding auxin biosynthetic enzymes does not differ between the generative organs of *L. luteus* that are shed and those that are not (except for one unigene in flowers). Closer look at the DEGs designed as involved in regulation of auxin availability (metabolism and conjugation) and transport suggest, that during organ abscission there is a huge rebuilding of auxin patterning. In falling flowers this is achieved by inhibition of synthesis, but also break-down and transport of this hormone. In pedicels of falling flowers the pattern probably changes by decreased auxin catabolism and changed transport. In falling pods auxin is probably less available because it is conjugated with proteins. Or it is rather released from conjugates in non-abscission pods.

In the flowers, we identified 11 DEGs involved in auxin signaling (Table 3 and Table S5). Flowers that remain on the plant over-express gene encoding key enzyme in IAA biosynthesis YUCCA8 and DAO, essential for auxin catabolism and maintenance of auxin homeostasis in reproductive organs (Zhao et al., 2013; Panoli et al., 2015). Analyses of auxin biosynthesis mutants (Cheng et al., 2006) indicate that localized IAA synthesis is critical for proper gynoecium morphogenesis in *A. thaliana* (for a review see: Hawkins and Liu, 2014). Up-regulation of YUCCA8 in FNAB suggests that IAA biosynthesis is also required for proper flower development in yellow lupine. Pollen grains of the *Arabidopsis dao* mutant did not germinate neither *in vitro* nor *in vivo* on the wild-type stigma, and no seeds were produced, which indicates that DAO plays a role in regulating anther dehiscence and pollen grain development (Zhao et al., 2013). This suggests that in falling lupine flowers, where DAO is less expressed, pollen grains germinate less effectively and in consequence fertilization does not occur, which is the cause of flower abscission. The gene encoding basipetal auxin transporter TORNADO 2 (TRN2) was up-regulated in FNAB. In *Arabidopsis* the auxin pattern resulting from TRN2 activity enables growth and organ organization by cell differentiation (Cnops et al., 2000; Cnops, 2006). Our finding indicate that flower abscission is correlated with inhibition of IAA biosynthesis in one area and simultaneous inhibition of

IAA break-down in the other area, and suppression of its transport. This may be caused by cues sent by older flowers, but further study is needed to verify this hypothesis and find possible countermeasures, to increase the number of maintained flowers on the plant. Among the many DEGs analyzed in the flower transcriptomes comparison, two DEGs interpreted as related to IAA signal transduction are worth mentioning: ARF4 and ARF7, up and down-regulated in FNAB, respectively (Table S5). An analysis of tomato *SLARF7* indicate that transcript level of this gene decreased after pollination and transgenic plants with decreased *SLARF7* expression formed parthenocarpic fruits (de Jong et al., 2009). In our RNA-Seq analysis, the ARF7 mRNA level was higher in the abscising flowers, thus supporting our hypothesis that abscising flowers of *L. luteus* are unfertilized. DEGs presenting similarity to ARF4 are present in all comparisons. In the flowers, ARF4 is over-expressed in FNAB. There is no clear evidence that ARF4 is involved in organ abscission. As genetic analyses indicate, ARF4 have functions in leaf and floral organ patterning and specify abaxial cell identity (Pekker et al., 2005). In tomato, *SLARF4* expression is high in flowers and young fruits, and decreases during fruit maturation and ripening (Sagar et al., 2013a,b). We suggest that ARF4 in yellow lupine could be associated with auxin regulation of development by way of symmetric cell division and organ polarity control.

Ethylene and auxin (IAA) are important regulators of abscission and the balance between these phytohormones determines where and when separation takes place (Ascough et al., 2005; Estornell et al., 2013). Most of the unigenes associated with auxin signaling found in this study were identified in the pedicels of flowers (44 out of 70) (Table 4 and Table S6). Two of these DEGs are responsible for auxin break-down: DAO, essential for IAA oxidation, and IAMT1, involved in formation of MeIAA. MeIAA demonstrates a different activity, as it more easily forms conjugates with proteins and carbohydrates and as a nonpolar molecule probably diffuses through membranes (Qin et al., 2005). Both DEGs were up-regulated in FPNAB. Three of the four DEGs encoding auxin transport proteins were up-regulated in FPNAB, which presumably resulted in a change in auxin distribution. LAX4 and LAX5, encoding the auxin influx carrier (Bainbridge et al., 2008; Petrášek and Friml, 2009; Vanneste and Friml, 2009) proteins were over-expressed in lupine FPNAB. Auxin efflux transporter BIG that controls elongation of pedicels and stem internodes through auxin action in *Arabidopsis* (Gil et al., 2001; Yamaguchi et al., 2007) showed a contrary tendency. Also, the expression of the gene encoding PINOID, a kinase that phosphorylates PINs thus leading to their internalization into cells (Petrášek and Friml, 2009), was reduced. We conclude that in FPNAB: (i) IAA break-down was reduced, (ii) the amount of its form that could diffuse out of cells was lower, (iii) auxin influx was enhanced and efflux was inhibited, (iv) and polar auxin transport was reduced. This is consistent with literature data that in poplar leaves prior to cell separation during abscission, an auxin maximum in AZ is formed (Jin et al., 2015) and that reduction of polar auxin transport increases sensitivity of the tissue to ET, which stimulates production of hydrolytic enzymes resulting in organ detachment (Estornell et al., 2013).

This indicates, that in yellow lupine the mechanism of organ abscission is similar to that of other plants.

The list of DEGs identified in the pedicels was rich in genes encoding other elements of auxin signal transduction: two encoded down- and up-regulated receptors (F-box protein SKP2A and auxin-binding protein ABP19a, respectively), three encoded down-regulated auxin response factors (ARF4, ARF7 and ARF19), and as many as 15 encoded up- or down-regulated AUX/IAA proteins (Table 4 and Table S6). In *Arabidopsis*, SKP2A is an F-box protein that binds auxin and connects its signaling with cell division (Jurado et al., 2010). Over-expression of SKP2A in pedicels of abscising flowers in yellow lupine is probably associated with the auxin-dependent regulation of cell division in the AZ. ABP receptors are known to regulate many processes. Upon auxin binding, they restrict internalization of PINs by inhibiting clathrin-mediated endocytosis (Robert et al., 2010; Grönes et al., 2015), thus they are involved in regulating polar auxin transport (Effendi et al., 2011). They also regulate the expression of genes encoding cell wall remodeling proteins, such as EXP, XTH and pectin (Paque et al., 2014). This suggests that during the growth of the pedicel of a flower remaining on the plant, ABP may be engaged in the regulation of cell growth and PIN-mediated auxin patterning. Among auxin-related DEGs in flower pedicels we also found ARF4 (one DEG up-, another down-regulated in FPNAB), ARF7 and ARF19 (both down-regulated in FPNAB). As mentioned before, ARF4 is probably involved in regulating polar cell division and cell wall modification. Genetic analysis of *Arabidopsis* mutants suggests that ARF7 and ARF19 are involved in abscission of floral organs (Estornell et al., 2013). In NGS data analysis of the abscission zone in tomato flowers and leaves ARF19 was also over-expressed in the AZ, while in soybean leaf abscission only ARF8 expression changes were observed (Kim et al., 2016). Our results indicate that ARF7 and ARF19 can be associated with flower abscission in lupine, but the mechanism of how these genes regulate abscission needs to be clarified.

In the falling pods we observed reduced amounts of ILR1-like 2 (IAA-amino acid hydrolase) transcripts responsible for IAA conjugation to amino acids (LeClere et al., 2002) and increased DEGs encoding auxin receptors such as TIR1, ABP19, and SKP2A (Table 5 and Table S7). This indicates an increase in free auxin concentrations and a growth of the sensitivity to IAA in these organs. The amount of transcripts encoding ARF2 and ARF4 in PNAB is higher than in PAB. It is suggested that ARF2 is a general repressor of auxin-regulated cell division (Ellis et al., 2005; Okushima et al., 2005) and plays an important role in growth of the ovule before fertilization that determines the final size of the seed (Schruff et al., 2005). Therefore, ARF4 and ARF2 may participate in the regulation of cell division in the developing pods of yellow lupine.

Ethylene

In the flower, we found only three DEGs involved in ethylene (ET) signal transduction, which suggests that in this organ ethylene did not play any primary role in abscission regulation (Table 3 and Table S5).

This was not the case, however, in the flower pedicels, as here 17 of the identified DEGs were involved in ethylene metabolism (Table 4 and Table S6). Three unigenes encoding the ET biosynthesis enzyme ACO1 were down-regulated in FPNAB. This confirms the outcome of study suggesting that in *L. luteus*, flowering abortion evoked by ABA resulted from its stimulatory effect on the expression of two main ET biosynthesis genes *LIACS* and *LIACO*, the ACC content and, consequently, the increase in that phytohormone production (Wilmowicz et al., 2016). As mentioned above, ethylene accumulation causes auxin transport inhibition and organ abscission. The same takes place in yellow lupine pedicels, as an increase in the expression of ET biosynthesis genes correlates with decrease in the expression of genes associated with auxin transport in the pedicels of abscising flowers. Additionally, fourteen genes encoding elements of ET signal transduction were up- or down-regulated in FPAB (Table 4 and Table S6). In RNA-Seq analysis of another plant species many genes related to different steps of the ethylene signaling transduction pathway were expressed in the tomato AZs both after flower removal (FAZ) and leaf deblading (LAZ) (Sundaresan et al., 2016). Also, *EIN3* and the genes from ERF family were over-expressed in the tomato flower and leaf AZs (Sundaresan et al., 2016) and in the FPAB of yellow lupine too.

In the pods, all the DEGs attributed to ethylene biosynthesis (ACS1 and ACS7) and signal transduction (ERF3 and ERF1A) were less expressed in PNAB, which indicated that in the abscising pods ET was accumulated in larger amounts (Table 5 and Table S7). Ethylene plays a major role in regulating the expression of cell wall-associated genes encoding PGs (Sitrit and Bennett, 1998; Hiwasa et al., 2004), EXPs (Rose et al., 1997; Hiwasa et al., 2003) and BGals (Karakurt and Huber, 2004; Nishiyama et al., 2007). Therefore, we suggest that an extensive up-regulation of cell wall-related DEGs in the abscising pods could, to some extent, be the result of an increased ethylene biosynthesis in these organs.

Gibberellin Biosynthesis Is Induced in Organs Maintained on Plant

Gibberellins are hormones involved in cell expansion, fruit set and growth (Katsumi and Ishida, 1991; Serrani et al., 2007). In contrast to auxin, GA biosynthesis was mainly altered in the tested samples. Changes also occurred at the level of signal transduction and catabolism, but to a minor extent.

The large number of identified genes related to gibberellin metabolism and signaling indicates the importance of GA in lupine flower fate. By comparing the gene expression between the flowers falling off and control, we found 7 DEGs involved in GA functioning. One of them, *GA2OX1* engaged in GA catabolism, was down-regulated, while the others—responsible for its biosynthesis (*GAO2*, *GA20OX2*, *GA20OX1*) and signal transduction (*GAI*, *GASA12*, and *GASA14*)—were up-regulated in FNAB (Table 3 and Table S5). The mRNA levels of GA biosynthesis genes rise after pollination in tomato (Rebers et al., 1999; Serrani et al., 2007). In addition, the transcript level of *SlGA2ox2* was found to be lower, that results in higher levels of GA in active form (Serrani et al., 2008). In *A. thaliana* and *P. sativum*, auxin probably acts as a signal from successfully

fertilized ovules that, in turn, stimulates GA biosynthesis and triggers fruit development (Ozga and Reinecke, 2003; Dorcey et al., 2009). Our RNA-Seq analysis shows that probably a similar scenario occurs in yellow lupine.

Similarly to the flowers, in the pedicels of lupine flowers GA biosynthesis was induced in FPNAB (Table 4 and Table S6). This suggests that in this part of non-abscising flowers the GA signaling pathway was also induced.

Our RNA-Seq analysis shows that GA plays an essential role in retaining the pods on the plant (Table 5 and Table S7). The transcription level of GA biosynthesis genes was already higher in FNAB, and the difference even rose in PNAB. All the 10 DEGs attributed to GA biosynthesis were strongly expressed in PNAB, including *GA20OX2* which was the most up-regulated DEG in the developing pods. At the same time, this up-regulation in PNAB means that in the abscising pods lower GA biosynthesis was ceasing their development. Interestingly, DEGs involved in GA signal transduction (mainly *GAMYB*) were significantly over-expressed in the pods being dropped. *GAMYB* is known as a regulator of flower induction (Gocal et al., 1999) and α -amylase activation in the aleurone layer of the seeds (Gómez-Cadenas et al., 2001; Kaneko et al., 2002). There is no evidence of their function in fruit abscission or development. Only one study demonstrates that in the seeds of *A. thaliana*, *GAMYB-like* genes are able to promote—but are not essential for—the progression of PCD in the aleurone layer (Alonso-Peral et al., 2010). GA also regulates exine formation and PCD of tapetal cells and the direct activation of *CYP703A3* by *GAMYB* is crucial for exine formation (Aya et al., 2009). Probably, *GAMYB* is involved in PCD induction during lupine pod abscission.

Involvement of Other Hormones in Lupine Organs Abscission

There is still little understanding of the participation of other plant hormones in cutting off generative organs. Our study aims to fill this gap.

Our RNA-Seq analysis revealed that ABA receptor *PYL2* was over-expressed in the abscising flowers and flower pedicels (Tables S5, S6). In the flower pedicels, other DEGs recognized as involved in abscisic acid signaling, such as *SAPK2*, *SAPK8*, and *PP2CA*, were also down-regulated in FPNAB. Simultaneously, *CYP707A4* engaged in ABA break-down was up-regulated. These data suggested that in the abscising flowers and flower pedicels ABA signaling pathways were being activated. In the pods of *L. luteus*, ABA catabolism gene *CYP707A1* was up-regulated in PNAB, which means that in the abscising pods this gene was less expressed, and ABA accumulation occurred. In apple, before fruits abscission a chronological increase in the level of ABA and ACC also takes place (Gómez-Cadenas et al., 2000). Additionally, *ABI3* involved in ABA signal transduction was more expressed in PNAB. *ABI3* has been demonstrated to directly induce the expression of storage protein genes (Ezcurra et al., 1999; Reidt et al., 2000; Kroj et al., 2003; Braybrook et al., 2006; Wang and Perry, 2013).

Among the DEGs in the flowers, we identified unigenes that showed homology to jasmonic acid (JA) and ester methyl jasmonate (JaMe) biosynthesis genes (Table S5). A similar

situation took place in the pods. DEGs connected with JA biosynthesis were all up-regulated in PNAB (Table 5 and Table S7), which suggests stimulation of metabolism of these hormones in the pods that were continuing to develop. In contrast, in the pedicels of non-abscising flowers, JA biosynthesis was decreased (Table 4 and Table S6). Mutations in JA biosynthesis genes in *A. thaliana* causes male sterility due to delayed anther development and shortened filaments (reviews in Wasternack, 2007; Browse, 2009). Additionally, it was shown that JA is also required for proper development of tomato embryo (Goetz et al., 2012). This again confirms our hypothesis that flowers undergoing abscission are not fertilized and in falling pods embryo development is stopped.

Cytokinins are involved in the regulation of cell division and expansion (Skoog et al., 1965) and CK effect on abscission is thought to be mediated by ethylene (Sipes and Einset, 1983; Grossmann, 1991; Cin et al., 2007). In the pedicels of non-abscising flowers (Table 4 and Table S6), more expressed were those DEGs that showed similarity to: CK receptor *AHK4*, two genes encoding cytokinin biosynthesis enzymes *LOG1* and *LOG3*, and two genes *CKX5* and *CKX3* encoding enzymes that catalyze CK oxidation to biologically inactive forms (Galuszka et al., 2007). In the pedicels of flowers undergoing abscission two other forms, namely *CKX1* and *CKX9*, were up-regulated. In *Arabidopsis* roots, CKX proteins show diverse subcellular and tissue-specific localization that suggests specific developmental and physiological functions of each gene (Werner et al., 2003). Opposite changes in the expression of CKX genes in the flower pedicels of *L. luteus* may be associated with the fact, that different processes regulated by CK occur in various cell compartments and tissues but more study is need to clarify the role of CK in organ abscission in lupine.

Brassinosteroids are endogenous plant hormones essential for the proper regulation of multiple physiological processes required for proper plant growth and development (Clouse and Sasse, 1998; Krishna, 2003; Sasse, 2003; Clouse, 2011). In FPNAB, 8 DEGs were recognized as involved in BR signaling, which suggests that BRs play an important role in the pedicels of the developing flowers (Table 4 and Table S6). Three DEGs involved in signal transduction (*Cyclin-D3*) were up-regulated in FPNAB, which indicates more intensive cell division in the growing pedicels. DEGs putatively encoding BR biosynthesis enzymes were up- and others down-regulated. BR biosynthesis can be conducted via many alternative pathways which interweave at many nodes and form a sort of metabolic grid (see KEGG map00905; Shimada et al., 2001). In the flower pedicels, some alternative routes are up-regulated, which probably leads to changes not only in BR concentrations, but also in BR composition. BR promote cell expansion through regulation of expression of the genes involved in cell wall modifications, cellulose biosynthesis, ion and water transport, and cytoskeleton rearrangements (Clouse and Sasse, 1998; Schumacher et al., 1999; Morillon et al., 2001; Vert et al., 2005; Kim and Wang, 2010). These findings correlate with our RNA-Seq analysis results; apart from many up-regulated cell wall-related DEGs, 12 DEGs showing similarity to aquaporins were also up-regulated in FPNAB (Table S6).

In the flower pedicels and pods, we identified five and six DEGs (indicating similarity to *TGA*, *SABP2*, and *PR1*) interpreted as SA-related (Tables 4, 5, Tables S6, S7), respectively. This situation is common during organ abscission. For example, in tomato, transcription factor *TGA* was expressed in FAZ and LAZ (Sundaresan et al., 2016). The transcription of several genes coding for pathogenesis-related (PR) proteins increased concomitantly with the onset of flower abscission of *Sambucus nigra* (Coupe et al., 1997).

The expression of the *DAD2* gene (another name: *DR14/DWARF14*) encoding the receptor of SL is differential in the flowers and pods (Tables 3, 5, Tables S5, S7). DR14 protein could work as an intercellular signaling molecule to fine-tune SL function (Kameoka et al., 2016). Recently, a hypothesis in which SL may indirectly influence seed size has been created. SL mutants show enhanced shoot branching and delayed leaf senescence, so they may have reduced nutrient remobilization resulting in a reduction in seed production. These findings indicate that SLs may affect the grain crop yield through leaf senescence and shoot branching regulation (Yamada and Umehara, 2015). We suggest that the increased transcription of the *DAD2* gene in the non-abscising flowers and pods can be linked to the nutrients mobilization.

Other Factors Probably Involved in Generative Organ Abscission of Yellow Lupine

Transcription Factors Involved in Generative Organ Abscission

Some transcription factors (TFs) characteristic for soybean leaf and flower abscission processes, such as *NAC*, *WRKY*, and *AIL/PLETHORA (PLT)* (Kim et al., 2016), were also up-regulated in FPAB of lupine (Table S9). Among the DEGs, there were *AIL/PLT* (3 and 2 unigenes respectively), *NAC* (29) and *WRKY* (12). For example, DEG c60377_g2 with $\log_2FC = -8.08$, a putative ortholog of *NAC* transcription factor 29, also called *NAP (NAC-LIKE, ACTIVATED BY AP3/PI)*, is closely associated with the senescence process of *Arabidopsis* rosette leaves and, possibly, in other plant species (Guo and Gan, 2006; Zhang and Gan, 2012). The expression of *ANAC019*, *ANAC055*, and *ANAC072* was induced by drought, high salinity, and abscisic acid (Tran et al., 2004). *NAC055* and *NAC072* were also over-expressed in the FPAB of lupine (Table S9). These results are consistent with previous studies showing up-regulation of this type of TF genes in the flower AZ of tomato (Sundaresan et al., 2016), and the fruit AZ during abscission of mature melon and olive fruit (Corbacho et al., 2013; Gil-Amado and Gomez-Jimenez, 2013). Another transcription factor, namely *MYB108*, was up-regulated in FAB, as well. This protein contributes to the regulation of stamen maturation and male fertility in response to jasmonate signaling in *A. thaliana* (Mandaokar and Browse, 2008).

Some TFs were more expressed in organs continuing development. For example, the bHLH TF, *MYC4* is up-regulated in PNAB (Table S9). The maximum expression of *MYC2*, *MYC3*, and *MYC4* in *A. thaliana* coincided with the developmental stages during which seed storage reserves

accumulated, suggesting these genes may affect the accumulation of seed storage reserves (Gao et al., 2016). Over-expression of *MYC4* in the developing pods of yellow lupine suggests that storage reserves accumulated in seeds.

PCD and ROS Are Associated with Abscission Process

Organs that completed the abscission, undergo changes that resemble the canonical plant PCD (van Doorn et al., 2011), among others, the stimulation of nucleases and reactive oxygen species (ROS) (Farage-Barhom et al., 2008; Sakamoto et al., 2008a; Meir et al., 2010; Bar-Dror et al., 2011). In tomato, the overexpression of antiapoptotic proteins and inhibition of a PCD-associated ribonuclease trigger abscission (Lers et al., 2006; Bar-Dror et al., 2011), whereas in pepper, ROS inhibitors prevent abscission suppressing H_2O_2 production (Sakamoto et al., 2008b).

Our results confirm these findings. Among the DEGs from all the compared libraries we identified numerous genes encoding proteins directly involved in protection against reactive oxygen species, such as peroxidase, as well as enzymes belonging to the carotenoid biosynthesis pathway, such as *CCD4*, *ZDS*, and *beta-carotene hydroxylase* (Tables S8–S10). Also, 16 DEGs displaying similarity to various nucleases were more expressed in FPAB (Table S9). Additionally, one of the most intensively expressed DEGs in the flower pedicels with an active AZ showed similarity to the vacuolar processing enzyme *VPE*, an executor of plant PCD (Hatsugai et al., 2004). *VPE* is a cysteine protease that cleaves a peptide bond at the C-terminal side of asparagine and aspartic acid (Wang et al., 2009) and plays an essential role in the regulation of the lytic system of plants during the processes of defense and development (Hara-Nishimura et al., 2005).

Water Transport Plays an Important Role Especially in the Developing Organs

Our RNA-Seq analysis revealed that water transport played an important role especially in the developing organs. Aquaporin *TIP2;1* was one of the common DEGs for all our study libraries. Additionally, transcripts of 12 DEGs encoding aquaporins belonging to the tonoplast intrinsic protein (*TIP*) and plasma membrane intrinsic protein (*PIP*) gene families were more accumulated in the pedicels of non-abscising flowers (Table S9). There were also 6 (5 up-regulated) and 5 (4 down-regulated) DEGs presenting similarity to genes encoding aquaporins in the flowers and pods, respectively (Tables S8, S10). *TIP1* and *TIP2* are preferentially associated with the large lytic vacuoles and vacuoles accumulating vegetative storage proteins, respectively (Gattolin et al., 2009). Moreover, intensive expression of *TIP* and *PIP* genes during cell expansion has been observed in numerous plant species (e.g., O'Brien et al., 2002; Liu et al., 2008; Saito et al., 2015). Up-regulation of aquaporin genes in *L. luteus* FNAB and FPNAB is probably associated with cell expansion. On the other hand, in the pods, beside *TIP2;1*, other DEGs belong to the *NIP* aquaporin family (that are involved in transport of water, but also other molecules; Dean et al., 1999; Liu et al., 2003; Wallace and Roberts, 2005; Takano et al., 2006; Choi and Roberts, 2007; Mitani-Ueno

et al., 2011) were over-expressed in PAB (Table S10). This was probably associated with the processes of nutrient reutilization.

Calcium Signaling Pathway Plays an Important Role in Yellow Lupine Pod Abscission

Calcium (Ca^{2+}) is crucial for numerous biological functions. In addition to its key roles in ensuring cell wall and membrane system structural integrity, it has been shown to act as an intracellular regulator in many aspects of plant growth and development, including stress responses (White and Broadley, 2003; Ranty et al., 2006), cell division and elongation, and fruit growth (Hepler, 2005). Recent RNA-Seq analyses of apple fruit abscission revealed that Ca^{2+} deficiency due to the down-regulation of genes encoding transporters of this cation could be a signal for the degeneration of the lateral apple fruits (Ferrero et al., 2015). In contrast, our study the abscising pods demonstrated over-expression of 26 unigenes associated with calcium transport, homeostasis and response (Table S10). Among them there were: a putative calcium-transporting ATPase, a calcium-dependent protein kinase, a probable calcium-binding protein, a cation/calcium exchanger and a calmodulin-like protein. This suggests that the Ca^{2+} signaling pathway plays an important role in yellow lupine pod abscission, although there is a need for further research in order to clarify which physiological processes these changes are associated with.

Fertilization Probably Did Not Occur in Abscising Flowers

In addition to the cell-wall related, *MYB108* and other mentioned genes, there were more unigenes associated with pollination that were differentially expressed between the developing flowers and the abscising ones. In the abscising flowers, we detected more mRNA of *MS5*, a gene essential for male fertility, especially for microspora and pollen grain production (Glover et al., 1998), and involved in regulating cell division after male meiosis I and II to facilitate meiotic exit and transition to G1 (Ross et al., 1997; Bulankova et al., 2010) (Table S8). Another gene, *NFD4*, encoding a protein required for polar nuclear fusion during female gametophyte development and karyogamy during fertilization (Portereiko et al., 2006), was also less expressed in FAB. This confirms the hypothesis that shed flowers are unfertilized.

Analysis of DEGs Common for All Comparisons Confirms Known Data but Also Reveals *MIR169* Is Involved in Generative Organ Abscission

Our analyses of DEGs common for all comparisons allowed determination of which changes were universally associated with the processes of generative organ abscission. By that means we distinguished 28 unigenes with different expression in all of the libraries compared. 25 of them were down-regulated, 2 were up-regulated in all comparisons, and only one—mainly *TIP2;1*—was up-regulated in the FPNAB, unlike the other samples (Table 6, Table S11).

Among the common DEGs, there were representatives of factors involved in the already discussed issues such as cell wall reorganization (*EGL11*) and water transport (aquaporin *TIP2;1*).

The EG45-like domain containing protein (PNP-A) also has a systemic role in H_2O and solute homeostasis (Ludidi et al., 2002). Numerous common DEGs are involved in stress response, among them *ZAT12*, *HSF24*, *RCI2B*, *5INPP5A*, *SPX3*, *MYB4*, or *SIB2*. *ZAT12* play a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis* (Davletova et al., 2005). *SIB2*, *Arabidopsis* sigma factor binding proteins, is an activator of the WRKY33 transcription factor in plant defense (Lai et al., 2011). *RCI2B* and *RCI2A* are two developmentally- and stress-regulated cold-inducible genes of *Arabidopsis* encoding highly conserved hydrophobic proteins expressed during the first stages of seed development and germination, in vascular bundles, pollen, and guard cells (Medina et al., 2001).

The presence of the DEG encoding the putative calcium-transporting ATPase 13 (*ACA13*), the enzyme catalyzing the hydrolysis of ATP coupled with the translocation of calcium from the cytosol out of the cell or into organelles (Iwano et al., 2014), confirmed that calcium signaling plays an important role in organ abscission. Interestingly, among common DEGs we identified one showing similarity to *QKY* encoding a predicted membrane-bound protein. By analogy to animal proteins with related domain topology, *QKY* is speculated to be involved in Ca^{2+} -dependent signaling and membrane trafficking. *QKY* interacts with receptor-like kinase *STRUBBELIG* (*SUB*) at plasmodesm (PD) to promote tissue morphogenesis, for example integument initiation and outgrowth during ovule development (Fulton et al., 2009; Vaddepalli et al., 2014). Our RNA-Seq findings indicate that cell-cell communication *via* *QKY* was an important factor determining organ fate.

Surprisingly, none of the common DEGs were associated with hormone metabolism or signaling. This was probably due to the fact that although the significant changes in hormone-related genes occurred in all of the tested samples, they were specific to a particular organ. However, the DEGs from the same cluster but with others gene number (*c62886_g2*, *c62886_g7*, *c62886_g4*) indicating similarity to *ARF4* proteins were presented in all comparisons (Tables S8–S10).

Our analysis revealed similarity of the common DEG *c125095_g1* to the *miR169e* precursor of *Glycine max* (*gma-MIR169e*) (Figure 5). MicroRNAs (miRNA) are small noncoding RNAs that control development, stress responses and hormone signaling or metabolism, in both animals and plants, by post-transcriptional regulation of gene expression (Bartel, 2009; Voinnet, 2009). In plants, 21 nt miRNAs are processed from long stem-loop precursor RNAs. The mature miRNA is incorporated into the RNA induced silencing complex (RISC) (Schott et al., 2012) to guide post-transcriptional gene silencing (PTGS) of complementary mRNA by cleavage and/or inhibition of translation (Brodersen and Voinnet, 2009). A sequence analysis revealed that the transcript of unigene *c125095_g1* sequence forms the hairpin characteristic for pre-miRNAs, wherein the stem was the sequence of the mature molecule of *miR169e* (Figure 5), thus indicating that it may be a source of functional mature miRNA. *MIR169* has been identified in many plant species (Sunkar and Jagadeeswaran, 2008) and their isoforms are involved in regulation various processes during development (Gonzalez-Ibeas et al., 2011) and in response to biotic (Singh

et al., 2012) or abiotic stresses (Zhou et al., 2010; Licausi et al., 2011; Zhao et al., 2011). In plants, the main targets of miR169 are genes that encode the subunit A of Nuclear Factor Y (Rhoades et al., 2002). In plants, NF-Y TFs have been linked to development (Lotan et al., 1998; Combier et al., 2006; Wenkel et al., 2006), signaling (Warpeha et al., 2007), stress responses (Nelson et al., 2007; Li et al., 2008; Liu and Howell, 2010; Zhao et al., 2011), carbohydrate metabolism and cell wall modification (Leyva-González et al., 2012). Among yellow lupine DEGs no NF-YA transcripts were found, although this may be related to the fact that miR169 may regulate the expression of its target genes also by inhibiting translation, which cannot be detected using RNA-Seq. Further study is needed to identify the processes regulated by miR169 in yellow lupine.

Several DEGs that we identified, such as *ARF2*, *ARF4* or *GAMYBA* were also regulated by the regulatory sRNAs - miR390, ta-siARF (Marin et al., 2010) and miR159 (Alonso-Peral et al., 2010), respectively. These findings shed new light on possible regulation of organ abscission by non-coding RNAs.

RT-qPCR Validation of RNA-Seq Results

In order to validate the results obtained via RNA-Seq analyses, we assessed the trends in expression of 11 chosen genes which displayed significantly different transcription level in all of the compared transcriptomes, using RT-qPCR. The results of the expression analysis of these genes supported the validity of our RNA-Seq (**Figure S1**).

Metabolic Changes Associated with the Processes of Generative Organ Abscission in Lupine

KEGG is a highly integrated database providing information on the biological systems and their relationships at the molecular, cellular and organism levels, particularly via the KEGG pathway maps (Kanehisa et al., 2007). An analysis of KEGG metabolic pathways, which was performed using the identified DEGs, allows for a global analysis of the metabolic changes that occur during abscission, both in pedicels with an AZ and the abscising organs. The identified DEGs were grouped into 4 major pathways from the KEGG pathway database (**Figures S2A, S3A, S4A**): “metabolism,” “genetic information processing,” “environmental information processing” and “organismal systems.” As “metabolism” appeared to be one of the most significant and highly represented categories in our study we ran an in-depth analysis of it, which is presented in **Figures S2B, S3B, S4B**.

Our analysis of enriched KEGG pathways from all the comparisons indicates that some metabolic pathways were modulated in all the cases, while some seemed to be characteristic for specific plant organs. “Carbohydrate metabolism” and “amino acid metabolism” were common enriched pathway categories for the DEGs from all the comparisons made. This suggests that these metabolism categories were a major target of changes during the abscission of generative organs in yellow lupine. It is worth mentioning that to the “Pentose and glucuronate interconversions” pathway from the “carbohydrate metabolism” category belong i.e., cell-wall related PME and PAE. This explains

the path’s strong enrichment in all library comparisons, and especially in the pods (**Figures 6–8**).

The KEGG pathways which were enriched particularly in FNAB were “Terpenoid backbone biosynthesis” from the “Metabolism of terpenoids and polyketides” category (**Figure 6**). Terpenoids, also known as isoprenoids, are a large class of natural products consisting of isoprene (C5) units. In higher plants, isoprenoids participate in a wide variety of biological functions. Specific examples include photosynthetic pigments (chlorophylls and carotenoids) or hormones (ABA, GA, CK, and BR). The enrichment of these KEGG pathways in the flower could be associated with hormone biosynthesis. Other groups of up-regulated DEGs in FNAB belonged to the “Genetic Information Processing” category, and particularly to the “Replication and repair” and “Translation” pathways (38 DEGs) (**Figure 6** and **Figure S2A**). The fact that, apart from these pathways, amino acid biosynthesis and metabolism and nucleotide metabolism pathways were enriched, too, indicates that cell division and intensive *de novo* protein synthesis were induced in FNAB or from the other side these processes are inhibited in FAB.

Other pathways from the KEGG metabolism category in which DEGs identified in flower pedicels were indexed fell under the “Environmental Information Processing” heading, and included the “Plant hormone signal transduction pathway,” which confirmed the importance of changes in hormonal balance in the studied pedicels (**Figure 7**). This could be associated with the “Inositol phosphate metabolism” pathway that was over-represented in FPAB. We found that DEGs showing similarity to, *inter alia*, non-specific phospholipase C (NPC), belonged to this pathway. NPCs play important roles in many processes as phospholipid-to-galactosyl DAG exchange, growth and development associated with hormone signaling and stress responses (reviewed in: Pokotylo et al., 2013). The same phospholipases were included in the “Ether lipid metabolism” and “Glycerophospholipid metabolism” pathways that were strongly up-regulated in the abscising flower pedicels. In contrast, the “Biosynthesis of unsaturated fatty acid” and “Fatty acid metabolism” pathways that are probably associated with cell membrane functioning were enriched in the flowers that continued to develop.

In the pods, the “Plant-pathogen interaction” pathway from the “organismal systems” major category was strongly enriched due to high numbers of up-regulated DEGs related to PR1, calcium-related genes and WRKY in PAB (**Figure 8**). Another pathway enriched in the abscising pods was “Phosphatidylinositol signaling system,” which include DEGs encoding calmodulines and phospholipase A and C. In PAB, within the “amino acid metabolism” category we found DEGs showing similarity to enzymes involved in protein degradation, such as Cysteine proteinase (e.g., c134689_g1, c73956_g1, **Table S10**) (Martínez et al., 2012) or Proline dehydrogenase (e.g., c60845_g3, c53844_g2, **Table S10**) (Funck et al., 2010; Monteoliva et al., 2014), which confirmed our DEGs analysis indicating a strong induction of degradation processes in PAB. Only one pathway, “diterpenoid biosynthesis” related to GA

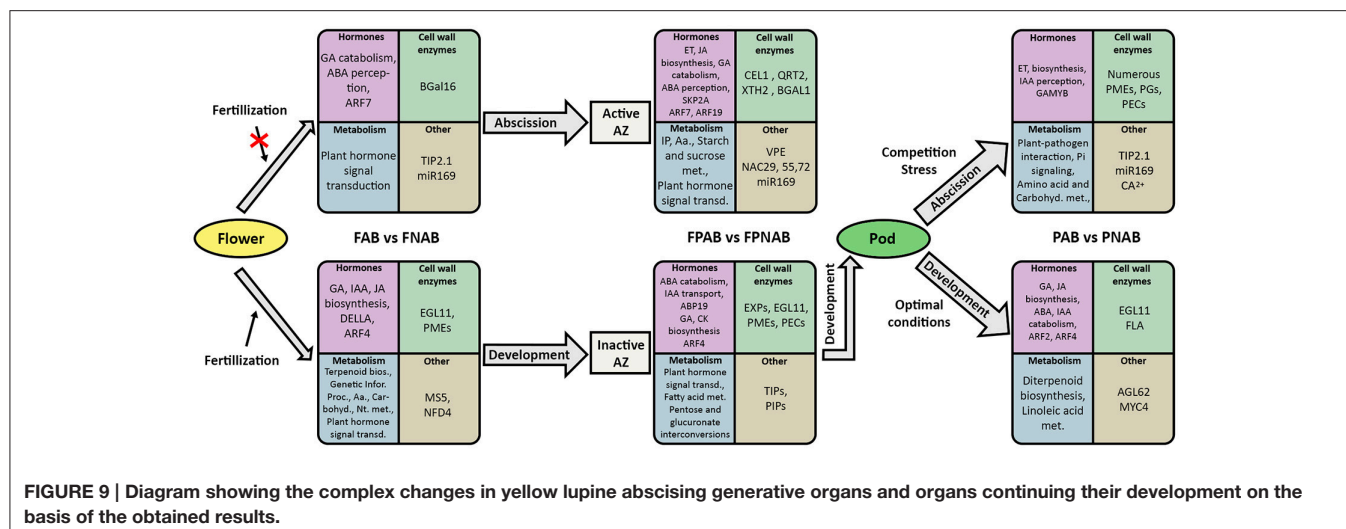


FIGURE 9 | Diagram showing the complex changes in yellow lupine abscising generative organs and organs continuing their development on the basis of the obtained results.

biosynthesis, was clearly enriched in PNAB, which fits well to our analysis of hormone-related DEGs in the pods.

CONCLUSIONS AND FUTURE PERSPECTIVES

In our work we analyzed global gene expression in the flowers, flower pedicels, and pods collected from *L. luteus* with the purpose of elucidating the molecular mechanisms and metabolic pathways involved in physiological abscission, and assessing the role of generative organs in this process.

As it might be expected, the process is regulated comprehensively and leads to the modification of gene expression, consequently causing hormonal, metabolic, and structural changes. The expression of certain genes changes in all of the compared libraries, while that of others is modified specifically in particular organs.

We propose a model, in which (Figure 9), unfertilized flowers of *L. luteus* induce the process of falling off and the activation of the AZ in the flower pedicel. In the fertilized flowers, changes related to further development and pod formation occur. The pedicels of these flowers also undergo structural modifications associated with the transport of water and nutrients and the maintenance of the growing fruit. The fate of the formed pods depends upon the sequence of establishing and the environmental conditions. Under unfavorable conditions growth is arrested, nutrient reutilization processes are activated, and cell wall degradation and PCD processes are launched. In all the cases, changes occurred in respect to the expression of a number of genes associated with the functioning of the cell walls, hormones and the metabolism of sugars and amino acids. Also, differences in the metabolism of hormones in the pedicels, as well as the signaling system and stress response in the pods, were observed. The observed increased *MIR169* expression in all of the abscising organs was a new discovery.

This study provides promising advances in our understanding of generative organ abscission that may ultimately lead to

improved control of *L. luteus* flower and pod set. In order to establish a more precise model and identify the abscission signal trigger, further studies are necessary. Further research is needed also to resolve whether here described changes in generative organs initiate before or after AZ activation, and whether they are the cause or result of abscission. Nevertheless, our study represents an important step in elucidating the biological pathways engaged in generative organ abscission, and more research is being conducted in order to unravel new genes and functions involved in this process.

AUTHOR CONTRIBUTIONS

PG conception and design of the experiment, plant and RNA sample preparation, data analysis, interpretation of data and wrote the manuscript; WW, KM, NK, and JKe conception and design of the experiment, plant and RNA sample preparation; WG and MK data analysis and manuscript preparation; JKo conception of the experiment and interpretation of data for the work.

FUNDING

This research was funded by the Polish Ministry of Agriculture and Rural Development grant No. 149/2011, the program supported by Resolution of the Council of Ministers (RM-111-222-15) in association with the Institute of Plant Genetics, Polish Academy of Sciences, and by The National Science Centre SONATA grant No. 2015/19/D/NZ9/03601 and NCN grant No. 2011/01/B/NZ9/03819.

ACKNOWLEDGMENTS

We would like to thank Michal Szczesniak from Ideas4biology Sp. z o. o. for performing the RNA sequencing data analysis and consulting.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00641/full#supplementary-material>

Figure S1 | qPCR validation of differential expression patterns of selected unigenes. The relative quantification of the gene expression level was determined with qPCR assay using *ACT1N* as a reference gene. The values presented on top of each bar indicate the expression level derived from the RNA-Seq data.

Figure S2 | Annotation of *L. luteus* DEGs from an FAB vs. FNAB library comparison by KEGG database. (A) Distribution of DEGs into KEGG biological categories. The numbers of KEGG pathways belonging to a category are shown in brackets. (B) Classification of DEGs into the KEGG “Metabolism” category. The numbers of DEGs belonging to a pathway are shown.

Figure S3 | Annotation of *L. luteus* DEGs from an FPAB vs. FPNAB library comparison by KEGG database. (A) Distribution of DEGs into KEGG biological categories. The numbers of KEGG pathways belonging to a category are shown in brackets. (B) Classification of DEGs into the KEGG “Metabolism” category. The numbers of DEGs belonging to a pathway are shown.

Figure S4 | Annotation of *L. luteus* DEGs from comparison of PAB vs. PNAB libraries by KEGG database. (A) Distribution of DEGs into KEGG biological categories. The numbers of KEGG pathways belonging to a category are shown in brackets. (B) Classification of DEGs into the KEGG “Metabolism” category. The numbers of DEGs belonging to a pathway are shown in brackets.

Table S1 | List of primers used for the RT-qPCR validation and the RT-qPCR validation results.

Table S2 | Quality filtering and statistics of raw reads obtained for eight transcriptome libraries of *L. luteus*.

Table S3 | Statistical analysis of differentially expressed gene in compared libraries of *L. luteus*.

Table S4 | Differential expression patterns of cell wall-related genes in library comparison of FAB vs. FNAB, FPAB vs. FPNAB, PAB vs. PNAB.

Table S5 | Differential expression patterns of plant hormone signaling-related genes in library comparison of FAB vs. FNAB.

Table S6 | Differential expression patterns of plant hormone signaling-related genes in library comparison of FPAB vs. FPNAB.

Table S7 | Differential expression patterns of plant hormone signaling-related genes in library comparison of PAB vs. PNAB.

Table S8 | List and annotation of differentially expressed genes (DEGs) between FAB and FNAB with PPDE < 0.05.

Table S9 | List and annotation of differentially expressed genes (DEGs) between FPAB and FPNAB with PPDE < 0.05.

Table S10 | List and annotation of differentially expressed genes (DEGs) between PAB and PNAB with PPDE < 0.05.

Table S11 | List and annotation of common DEGs identified in all three library comparisons of FAB vs. FNAB, FPAB vs. FPNAB, PAB vs. PNAB with FPKM (Fragment Per Kilobase of exon per Million fragments mapped) in each of the libraries.

Table S12 | The enriched pathways (KEGG) of DEGs from library comparison of FAB vs. FNAB.

Table S13 | The enriched pathways (KEGG) of DEGs from library comparison of FPAB vs. FPNAB.

Table S14 | The enriched pathways (KEGG) of DEGs from library comparison of PAB vs. PNAB.

REFERENCES

- Alonso-Peral, M. M., Li, J., Li, Y., Allen, R. S., Schnippenkoetter, W., Ohms, S., et al. (2010). The microRNA159-regulated *GAMYB*-like genes inhibit growth and promote programmed cell death in *Arabidopsis*. *Plant Physiol.* 154, 757–771. doi: 10.1104/pp.110.160630
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Arteca, R. N. (1996). *Plant Growth Substances: Principles and Applications*. New York, NY: Springer.
- Ascough, G. D., Nogemane, N., Mtshali, N. P., van Staden, J., and Bornman, C. H. (2005). Flower abscission: environmental control, internal regulation and physiological responses of plants. *S. Afr. J. Bot.* 71, 287–301. doi: 10.1016/S0254-6299(15)30101-0
- Aya, K., Ueguchi-Tanaka, M., Kondo, M., Hamada, K., Yano, K., Nishimura, M., et al. (2009). Gibberellin modulates anther development in rice via the transcriptional regulation of *GAMYB*. *Plant Cell Online* 21, 1453–1472. doi: 10.1105/tpc.108.062935
- Bainbridge, K., Guyomarc’h, S., Bayer, E., Swarup, R., Bennett, M., Mandel, T., et al. (2008). Auxin influx carriers stabilize phyllotactic patterning. *Genes Dev.* 22, 810–823. doi: 10.1101/gad.462608
- Bar-Dror, T., Dermastia, M., Kladnik, A., Žnidarič, M. T., Novak, M. P., Meir, S., et al. (2011). Programmed cell death occurs asymmetrically during abscission in tomato. *Plant Cell* 23, 4146–4163. doi: 10.1105/tpc.111.092494
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Belfield, E. J., Ruperti, B., Roberts, J. A., and McQueen-Mason, S. (2005). Changes in expansin activity and gene expression during ethylene-promoted leaflet abscission in *Sambucus nigra*. *J. Exp. Bot.* 56, 817–823. doi: 10.1093/jxb/eri076
- Botton, A., Eccher, G., Forcato, C., Ferrarini, A., Begheldo, M., Zermiani, M., et al. (2011). Signaling pathways mediating the induction of apple fruitlet abscission. *Plant Physiol.* 155, 185–208. doi: 10.1104/pp.110.165779
- Braybrook, S. A., Stone, S. L., Park, S., Bui, A. Q., Le, B. H., Fischer, R. L., et al. (2006). Genes directly regulated by *LEAFY COTYLEDON2* provide insight into the control of embryo maturation and somatic embryogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3468–3473. doi: 10.1073/pnas.0511331103
- Brodersen, P., and Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. *Nat. Rev. Mol. Cell Biol.* 10, 141–148. doi: 10.1038/nrm2619
- Browse, J. (2009). The power of mutants for investigating jasmonate biosynthesis and signaling. *Phytochemistry* 70, 1539–1546. doi: 10.1016/j.phytochem.2009.08.004
- Brummell, D. A., Dal Cin, V., Crisosto, C. H., and Labavitch, J. M. (2004). Cell wall metabolism during maturation, ripening and senescence of Peach fruit. *J. Exp. Bot.* 55, 2029–2039. doi: 10.1093/jxb/erh227
- Bulankova, P., Riehs-Kearnan, N., Nowack, M. K., Schnittger, A., and Riha, K. (2010). Meiotic progression in *Arabidopsis* is governed by complex regulatory interactions between SMG7, TDM1, and the meiosis I-specific cyclin TAM. *Plant Cell* 22, 3791–3803. doi: 10.1105/tpc.110.078378
- Cheng, Y., Dai, X., and Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev.* 20, 1790–1799. doi: 10.1101/gad.1415106
- Cho, H.-T., and Cosgrove, D. J. (2000). Altered expression of expansin modulates leaf growth and pedicel abscission in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 9783–9788. doi: 10.1073/pnas.160276997
- Choi, W.-G., and Roberts, D. M. (2007). *Arabidopsis* NIP2;1, a major intrinsic protein transporter of lactic acid induced by anoxic stress. *J. Biol. Chem.* 282, 24209–24218. doi: 10.1074/jbc.M700982200
- Cin, V. D., Boschetti, A., Dorigoni, A., and Ramina, A. (2007). Benzylaminopurine application on two different apple cultivars (*Malus domestica*) displays new and unexpected fruitlet abscission features. *Ann. Bot.* 99, 1195–1202. doi: 10.1093/aob/mcm062
- Clements, J., and Atkins, C. (2001). Characterization of a non-abscission mutant in *Lupinus angustifolius*. I. Genetic and structural aspects. *Am. J. Bot.* 88, 31–42.
- Clouse, S. D. (2011). Brassinosteroids. *Arab. B.* 9:e0151. doi: 10.1199/tab.0151

- Clouse, S. D., and Sasse, J. M. (1998). Brassinosteroids: essential regulators of plant growth and development. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 427–451. doi: 10.1146/annurev.arplant.49.1.427
- Cnops, G. (2006). The *TORNADO1* and *TORNADO2* genes function in several patterning processes during early leaf development in *Arabidopsis thaliana*. *Plant Cell Online* 18, 852–866. doi: 10.1105/tpc.105.040568
- Cnops, G., Wang, X., Linstead, P., Van Montagu, M., Van Lijsebettens, M., and Dolan, L. (2000). *TORNADO1* and *TORNADO2* are required for the specification of radial and circumferential pattern in the *Arabidopsis* root. *Development* 127, 3385–3394.
- Comber, J.-P., Frugier, F., de Billy, F., Boualem, A., El-Yahyaoui, F., Moreau, S., et al. (2006). MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes Dev.* 20, 3084–3088. doi: 10.1101/gad.402806
- Corbacho, J., Romojaro, F., Pech, J.-C., Latché, A., and Gomez-Jimenez, M. C. (2013). Transcriptomic events involved in Melon mature-fruit abscission comprise the sequential induction of cell-wall degrading genes coupled to a stimulation of endo and exocytosis. *PLoS ONE* 8:e58363. doi: 10.1371/journal.pone.0058363
- Coupe, S. A., Taylor, J. E., and Roberts, J. A. (1997). Temporal and spatial expression of mRNAs encoding pathogenesis-related proteins during ethylene-promoted leaflet abscission in *Sambucus nigra*. *Plant Cell Environ.* 20, 1517–1524. doi: 10.1046/j.1365-3040.1997.d01-54.x
- Davletova, S., Schlauch, K., Coutu, J., and Mittler, R. (2005). The zinc-finger protein Zat12 plays a central role in reactive oxygen and abiotic stress signaling in *Arabidopsis*. *Plant Physiol.* 139, 847–856. doi: 10.1104/pp.105.068254
- de Jong, M., Wolters-Arts, M., Feron, R., Mariani, C., and Vriezen, W. H. (2009). The *Solanum lycopersicum* Auxin Response Factor7 (SlARF7) regulates auxin signaling during tomato fruit set and development. *Plant J.* 57, 160–170. doi: 10.1111/j.1365-313X.2008.03671.x
- de Jong, M., Wolters-Arts, M., García-Martínez, J. L., Mariani, C., and Vriezen, W. H. (2011). The *Solanum lycopersicum* Auxin Response Factor7 (SlARF7) mediates cross-talk between auxin and gibberellin signalling during tomato fruit set and development. *J. Exp. Bot.* 62, 617–626. doi: 10.1093/jxb/erq293
- Dean, R. M., Rivers, R. L., Zeidel, M. L., and Roberts, D. M. (1999). Purification and functional reconstitution of soybean nodulin 26 an aquaporin with water and glycerol transport properties. *Biochemistry* 38, 347–353. doi: 10.1021/bi982110c
- del Campillo, E., and Bennett, A. B. (1996). Pedicel breakstrength and cellulase gene expression during tomato flower abscission. *Plant Physiol.* 111, 813–820.
- del Campillo, E., Durbin, M., and Lewis, L. N. (1988). Changes in two forms of membrane-associated cellulase during ethylene-induced abscission. *Plant Physiol.* 88, 904–909.
- Dokoozlian, N. K., and Peacock, W. L. (2001). Gibberellic acid applied at bloom reduces fruit set and improves size of “Crimson Seedless” table grapes. *HortScience* 36, 706–709.
- Dorsey, E., Urbez, C., Blázquez, M. A., Carbonell, J., and Perez-Amador, M. A. (2009). Fertilization-dependent auxin response in ovules triggers fruit development through the modulation of gibberellin metabolism in *Arabidopsis*. *Plant J.* 58, 318–332. doi: 10.1111/j.1365-313X.2008.03781.x
- Effendi, Y., Rietz, S., Fischer, U., and Scherer, G. F. E. (2011). The heterozygous *abp1/ABP1* insertion mutant has defects in functions requiring polar auxin transport and in regulation of early auxin-regulated genes. *Plant J.* 65, 282–294. doi: 10.1111/j.1365-313X.2010.04420.x
- Ellis, C. M., Nagpal, P., Young, J. C., Hagen, G., Guilfoyle, T. J., and Reed, J. W. (2005). Auxin Response Factor1 and Auxin Response Factor2 regulate senescence and floral organ abscission in *Arabidopsis thaliana*. *Development* 132, 4563–4574. doi: 10.1242/dev.02012
- Estornell, L. H., Agustí, J., Merelo, P., Talón, M., and Tadeo, F. R. (2013). Elucidating mechanisms underlying organ abscission. *Plant Sci.* 199–200, 48–60. doi: 10.1016/j.plantsci.2012.10.008
- Ezcurra, I., Ellerström, M., Wycliffe, P., Stålberg, K., and Rask, L. (1999). Interaction between composite elements in the napA promoter: both the B-box ABA-responsive complex and the RY/G complex are necessary for seed-specific expression. *Plant Mol. Biol.* 40, 699–709.
- Farage-Barhom, S., Burd, S., Sonogo, L., Perl-Treves, R., and Lers, A. (2008). Expression analysis of the *BFN1* nuclease gene promoter during senescence, abscission, and programmed cell death-related processes. *J. Exp. Bot.* 59, 3247–3258. doi: 10.1093/jxb/ern176
- Ferrero, S., Carretero-Paulet, L., Mendes, M. A., Botton, A., Eccher, G., Masiero, S., et al. (2015). Transcriptomic signatures in seeds of apple (*Malus domestica* L. Borkh) during fruitlet abscission. *PLoS ONE* 10:e0120503. doi: 10.1371/journal.pone.0120503
- Fragkostefanakis, S., Dandachi, F., and Kalaitzis, P. (2012). Expression of arabinogalactan proteins during tomato fruit ripening and in response to mechanical wounding, hypoxia and anoxia. *Plant Physiol. Biochem.* 52, 112–118. doi: 10.1016/j.plaphy.2011.12.001
- Frankowski, K., Wilmowicz, E., Kućko, A., Zienkiewicz, A., Zienkiewicz, K., and Kopcewicz, J. (2015). Profiling the *BLADE-ON-PETIOLE* gene expression in the abscission zone of generative organs in *Lupinus luteus*. *Acta Physiol. Plant* 37:220. doi: 10.1007/s11738-015-1972-y
- Fulton, L., Batoux, M., Vaddepalli, P., Yadav, R. K., Busch, W., Andersen, S. U., et al. (2009). *DETORQUEO*, *QUIRKY*, and *ZERZAUST* represent novel components involved in organ development mediated by the receptor-like kinase STRUBBELIG in *Arabidopsis thaliana*. *PLoS Genet.* 5:e1000355. doi: 10.1371/journal.pgen.1000355
- Funck, D., Eckard, S., and Müller, G. (2010). Non-redundant functions of two proline dehydrogenase isoforms in *Arabidopsis*. *BMC Plant Biol.* 10:70. doi: 10.1186/1471-2229-10-70
- Galuszka, P., Popelková, H., Werner, T., Frébortová, J., Pospíšilová, H., Milk, V., et al. (2007). Biochemical characterization of cytokinin oxidases/dehydrogenases from *Arabidopsis thaliana* expressed in *Nicotiana tabacum* L. *J. Plant Growth Regul.* 26, 255–267. doi: 10.1007/s00344-007-9008-5
- Gao, C., Qi, S., Liu, K., Li, D., Jin, C., Li, Z., et al. (2016). MYC2, MYC3, and MYC4 function redundantly in seed storage protein accumulation in *Arabidopsis*. *Plant Physiol. Biochem.* 108, 63–70. doi: 10.1016/j.plaphy.2016.07.004
- Gattolin, S., Sorieul, M., Hunter, P. R., Khonsari, R. H., and Frigerio, L. (2009). *In vivo* imaging of the tonoplast intrinsic protein family in *Arabidopsis* roots. *BMC Plant Biol.* 9:133. doi: 10.1186/1471-2229-9-133
- Gil, P., Dewey, E., Friml, J., Zhao, Y., Snowden, K. C., Putterill, J., et al. (2001). BIG: a calossin-like protein required for polar auxin transport in *Arabidopsis*. *Genes Dev.* 15, 1985–1997. doi: 10.1101/gad.905201
- Gil-Amado, J. A., and Gomez-Jimenez, M. C. (2013). Transcriptome analysis of mature fruit abscission control in Olive. *Plant Cell Physiol.* 54, 244–269. doi: 10.1093/pcp/pcs179
- Giné-Bordonaba, J., Echeverría, G., Ubach, D., Aguiló-Aguayo, I., López, M. L., and Larrigaudière, C. (2017). Biochemical and physiological changes during fruit development and ripening of two sweet Cherry varieties with different levels of cracking tolerance. *Plant Physiol. Biochem.* 111, 216–225. doi: 10.1016/j.plaphy.2016.12.002
- Glover, J., Grelon, M., Craig, S., Chaudhury, A., and Dennis, E. (1998). Cloning and characterization of *MS5* from *Arabidopsis*: a gene critical in male meiosis. *Plant J.* 15, 345–356.
- Gocal, G. F., Poole, A. T., Gubler, F., Watts, R. J., Blundell, C., and King, R. W. (1999). Long-day up-regulation of a *GAMYB* gene during *Lolium temulentum* inflorescence formation. *Plant Physiol.* 119, 1271–1278. doi: 10.1104/PP.119.4.1271
- Goetz, S., Hellwege, A., Stenzel, I., Kutter, C., Hauptmann, V., Forner, S., et al. (2012). Role of cis-12-oxo-phytodienoic acid in tomato embryo development. *Plant Physiol.* 158, 1715–1727. doi: 10.1104/pp.111.192658
- Gómez-Cadenas, A., Mehouchi, J., Tadeo, F. R., Primo-Millo, E., and Talon, M. (2000). Hormonal regulation of fruitlet abscission induced by carbohydrate shortage in citrus. *Planta* 210, 636–643. doi: 10.1007/s004250050054
- Gómez-Cadenas, A., Zentella, R., Walker-Simmons, M. K., and Ho, T. H. (2001). Gibberellin/abscisic acid antagonism in barley aleurone cells: site of action of the protein kinase PKABA1 in relation to gibberellin signaling molecules. *Plant Cell* 13, 667–679. doi: 10.1105/tpc.13.3.667
- González-Carranza, Z. H., Whitelaw, C. A., Swarup, R., and Roberts, J. A. (2002). Temporal and spatial expression of a polygalacturonase during leaf and flower abscission in oilseed rape and *Arabidopsis*. *Plant Physiol.* 128, 534–543. doi: 10.1104/pp.010610
- Gonzalez-Ibeas, D., Blanca, J., Donaire, L., Saladié, M., Mascarell-Creus, A., Cano-Delgado, A., et al. (2011). Analysis of the melon (*Cucumis melo*) small RNAome by high-throughput pyrosequencing. *BMC Genomics* 12:393. doi: 10.1186/1471-2164-12-393
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq

- data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Griffiths-Jones, S. (2004). The microRNA registry. *Nucleic Acids Res.* 32, D109–D111. doi: 10.1093/nar/gkh023
- Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A., and Enright, A. J. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144. doi: 10.1093/nar/gkj112
- Griffiths-Jones, S., Saini, H. K., van Dongen, S., and Enright, A. J. (2008). miRBase: tools for microRNA genomics. *Nucleic Acids Res.* 36, D154–D158. doi: 10.1093/nar/gkm952
- Grones, P., Chen, X., Simon, S., Kaufmann, W. A., De Rycke, R., Nodzyński, T., et al. (2015). Auxin-binding pocket of ABP1 is crucial for its gain-of-function cellular and developmental roles. *J. Exp. Bot.* 66, 5055–5065. doi: 10.1093/jxb/erv177
- Grossmann, K. (1991). Induction of leaf abscission in cotton is a common effect of urea- and adenine-type cytokinins. *Plant Physiol.* 95, 234–237.
- Gunawardena, A. H. L. A. N., Greenwood, J. S., and Dengler, N. G. (2007). Cell wall degradation and modification during programmed cell death in lace plant, *Aponogeton madagascariensis* (Aponogetonaceae). *Am. J. Bot.* 94, 1116–1128. doi: 10.1093/ajb.94.7.1116
- Guo, Y., and Gan, S. (2006). AtNAP, a NAC family transcription factor, has an important role in leaf senescence. *Plant J.* 46, 601–612. doi: 10.1111/j.1365-3113X.2006.02723.x
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hamiaux, C., Drummond, R. S. M., Janssen, B. J., Ledger, S. E., Cooney, J. M., Newcomb, R. D., et al. (2012). DAD2 is an α/β hydrolase likely to be involved in the perception of the plant branching hormone, strigolactone. *Curr. Biol.* 22, 2032–2036. doi: 10.1016/j.cub.2012.08.007
- Han, Y., Ban, Q., Li, H., Hou, Y., Jin, M., Han, S., et al. (2016). DkXTH8, a novel xyloglucan endotransglucosylase/hydrolase in persimmon, alters cell wall structure and promotes leaf senescence and fruit postharvest softening. *Sci. Rep.* 6:39155. doi: 10.1038/srep39155
- Hara-Nishimura, I., Hatsugai, N., Nakaune, S., Kuroyanagi, M., and Nishimura, M. (2005). Vacuolar processing enzyme: an executor of plant cell death. *Curr. Opin. Plant Biol.* 8, 404–408. doi: 10.1016/j.pbi.2005.05.016
- Hatsugai, N., Kuroyanagi, M., Yamada, K., Meshi, T., Tsuda, S., Kondo, M., et al. (2004). A plant vacuolar protease, VPE, mediates virus-induced hypersensitive cell death. *Science* 305, 855–858. doi: 10.1126/science.1099859
- Hawkins, C., and Liu, Z. (2014). A model for an early role of auxin in *Arabidopsis* gynoecium morphogenesis. *Front. Plant Sci.* 5:327. doi: 10.3389/fpls.2014.00327
- Hepler, P. K. (2005). Calcium: a central regulator of plant growth and development. *Plant Cell* 17, 2142–2155. doi: 10.1105/tpc.105.032508
- Hiwasa, K., Nakano, R., Hashimoto, A., Matsuzaki, M., Murayama, H., Inaba, A., et al. (2004). European, Chinese and Japanese Pear fruits exhibit differential softening characteristics during ripening. *J. Exp. Bot.* 55, 2281–2290. doi: 10.1093/jxb/erh250
- Hiwasa, K., Rose, J. K. C., Nakano, R., Inaba, A., and Kubo, Y. (2003). Differential expression of seven α -expansin genes during growth and ripening of pear fruit. *Physiol. Plant.* 117, 564–572. doi: 10.1034/j.1399-3054.2003.00064.x
- Iwano, M., Igarashi, M., Tarutani, Y., Kaothien-Nakayama, P., Nakayama, H., Moriyama, H., et al. (2014). A pollen coat-inducible autoinhibited Ca^{2+} -ATPase expressed in stigmatic papilla cells is required for compatible pollination in the *Brassicaceae*. *Plant Cell* 26, 636–649. doi: 10.1105/tpc.113.121350
- Jin, X., Zimmermann, J., Polle, A., and Fischer, U. (2015). Auxin is a long-range signal that acts independently of ethylene signaling on leaf abscission in *Populus*. *Front. Plant Sci.* 6:634. doi: 10.3389/fpls.2015.00634
- Jinn, T. L., Stone, J. M., and Walker, J. C. (2000). HAESA, an *Arabidopsis* leucine-rich repeat receptor kinase, controls floral organ abscission. *Genes Dev.* 14, 108–117. doi: 10.1101/gad.14.1.108
- Jurado, S., Abraham, Z., Manzano, C., López-Torrejón, G., Pacios, L. F., and Del Pozo, J. C. (2010). The *Arabidopsis* cell cycle F-Box protein SKP2A binds to auxin. *Plant Cell* 22, 3891–3904. doi: 10.1105/tpc.110.078972
- Kameoka, H., Dun, E. A., Lopez-Obando, M., Brewer, P. B., de Saint Germain, A., Rameau, C., et al. (2016). Phloem transport of the receptor DWARF14 protein is required for full function of strigolactones. *Plant Physiol.* 172, 1844–1852. doi: 10.1104/pp.16.01212
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2007). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–D462. doi: 10.1093/nar/gkv1070
- Kaneko, M., Itoh, H., Ueguchi-Tanaka, M., Ashikari, M., and Matsuoka, M. (2002). The α -amylase induction in endosperm during rice seed germination is caused by gibberellin synthesized in epithelium. *Plant Physiol.* 128, 1264–1270. doi: 10.1104/pp.010785
- Kang, I.-H., Steffen, J. G., Portereiko, M. F., Lloyd, A., and Drews, G. N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20, 635–647. doi: 10.1105/tpc.107.055137
- Karakurt, Y., and Huber, D. J. (2004). Ethylene-induced gene expression, enzyme activities, and water soaking in immature and ripe watermelon (*Citrullus lanatus*) fruit. *J. Plant Physiol.* 161, 381–388. doi: 10.1078/0176-1617-01221
- Katsumi, M., and Ishida, K. (1991). “The gibberellin control of cell elongation,” in *gibberellins*, eds N. Takahashi, B. O. Phinney, and J. MacMillan (New York, NY: Springer), 211–219.
- Kim, J. (2014). Four shades of detachment: regulation of floral organ abscission. *Plant Signal. Behav.* 9:e976154. doi: 10.4161/15592324.2014.976154
- Kim, J., Sundaresan, S., Philosoph-Hadas, S., Yang, R., Meir, S., and Tucker, M. L. (2015). Examination of the abscission-associated transcriptomes for Soybean, Tomato, and *Arabidopsis* highlights the conserved biosynthesis of an extensible extracellular matrix and boundary layer. *Front. Plant Sci.* 6:1109. doi: 10.3389/fpls.2015.01109
- Kim, J., Yang, J., Yang, R., Sicher, R. C., Chang, C., and Tucker, M. L. (2016). Transcriptome analysis of Soybean leaf abscission identifies transcriptional regulators of organ polarity and cell fate. *Front. Plant Sci.* 7:125. doi: 10.3389/fpls.2016.00125
- Kozomara, A., and Griffiths-Jones, S. (2011). miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* 39, D152–D157. doi: 10.1093/nar/gkq1027
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Kim, T.-W., and Wang, Z.-Y. (2010). Brassinosteroid signal transduction from receptor kinases to transcription factors. *Annu. Rev. Plant Biol.* 61, 681–704. doi: 10.1146/annurev.arplant.043008.092057
- Krishna, P. (2003). Brassinosteroid-mediated stress responses. *J. Plant Growth Regul.* 22, 289–297. doi: 10.1007/s00344-003-0058-z
- Kroj, T., Savino, G., Valon, C., Giraudat, J., and Parcy, F. (2003). Regulation of storage protein gene expression in *Arabidopsis*. *Development* 130, 6065–6073. doi: 10.1242/dev.00814
- Kuroha, T., Tokunaga, H., Kojima, M., Ueda, N., Ishida, T., Nagawa, S., et al. (2009). Functional analyses of LONELY GUY cytokinin-activating enzymes reveal the importance of the direct activation pathway in *Arabidopsis*. *Plant Cell* 21, 3152–3169. doi: 10.1105/tpc.109.068676
- Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J.-Q., et al. (2011). *Arabidopsis* sigma factor binding proteins are activators of the wrky33 transcription factor in plant defense. *Plant Cell* 23, 3824–3841. doi: 10.1105/tpc.111.090571
- Lashbrook, C. C., and Cai, S. (2008). Cell wall remodeling in *Arabidopsis* stamen abscission zones: temporal aspects of control inferred from transcriptional profiling. *Plant Signal. Behav.* 3, 733–736. doi: 10.4161/psb.3.9.6489
- Lashbrook, C. C., Gonzalez-Bosch, C., and Bennett, A. B. (1994). Two divergent endo-beta-1,4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers. *Plant Cell Online* 6, 1485–1493. doi: 10.1105/tpc.6.10.1485
- LeClere, S., Tellez, R., Rampey, R. A., Matsuda, S. P. T., and Bartel, B. (2002). Characterization of a family of IAA-amino acid conjugate hydrolases from *Arabidopsis*. *J. Biol. Chem.* 277, 20446–20452. doi: 10.1074/jbc.M111955200

- Leng, N., and Kendzierski, C. (2015). *EBSeq: An R Package for Differential Expression Analysis using RNA-Seq Data*. R package version 1.14.10.
- Lers, A., Sonego, L., Green, P. J., and Burd, S. (2006). Suppression of LX ribonuclease in tomato results in a delay of leaf senescence and abscission. *Plant Physiol.* 142, 710–721. doi: 10.1104/pp.106.080135
- Leyva-González, M. A., Ibarra-Laclette, E., Cruz-Ramírez, A., and Herrera-Estrella, L. (2012). Functional and transcriptome analysis reveals an acclimatization strategy for abiotic stress tolerance mediated by *Arabidopsis* NF-YA Family Members. *PLoS ONE* 7:e48138. doi: 10.1371/journal.pone.0048138
- Li, W.-X., Oono, Y., Zhu, J., He, X.-J., Wu, J.-M., Iida, K., et al. (2008). The *Arabidopsis* NFYA5 transcription factor is regulated transcriptionally and posttranscriptionally to promote drought resistance. *Plant Cell* 20, 2238–2251. doi: 10.1105/tpc.108.059444
- Licausi, F., Weits, D. A., Pant, B. D., Scheible, W.-R., Geigenberger, P., and van Dongen, J. T. (2011). Hypoxia responsive gene expression is mediated by various subsets of transcription factors and miRNAs that are determined by the actual oxygen availability. *New Phytol.* 190, 442–456. doi: 10.1111/j.1469-8137.2010.03451.x
- Liu, D., Tu, L., Wang, L., Li, Y., Zhu, L., and Zhang, X. (2008). Characterization and expression of plasma and tonoplast membrane aquaporins in elongating cotton fibers. *Plant Cell Rep.* 27, 1385–1394. doi: 10.1007/s00299-008-0545-6
- Liu, J. X., and Howell, S. H. (2010). bZIP28 and NF-Y transcription factors are activated by ER stress and assemble into a transcriptional complex to regulate stress response genes in *Arabidopsis*. *Plant Cell* 22, 782–796. doi: 10.1105/tpc.109.072173
- Liu, L.-H., Ludewig, U., Gassert, B., Frommer, W. B., and von Wirén, N. (2003). Urea transport by nitrogen-regulated Tonoplast intrinsic proteins in *Arabidopsis*. *Plant Physiol.* 133, 1220–1228. doi: 10.1104/pp.103.027409
- Lotan, T., Ohto, M., Yee, K. M., West, M. A., Lo, R., Kwong, R. W., et al. (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. *Cell* 93, 1195–1205.
- Ludidi, N. N., Heazlewood, J. L., Seoighe, C., Irving, H. R., and Gehring, C. A. (2002). Expansin-like molecules: novel functions derived from common domains. *J. Mol. Evol.* 54, 587–594. doi: 10.1007/s00239-001-0055-4
- Mandaokar, A., and Browse, J. (2008). MYB108 acts together with MYB24 to regulate Jasmonate-mediated stamen maturation in *Arabidopsis*. *Plant Physiol.* 149, 851–862. doi: 10.1104/pp.108.132597
- Marin, E., Jouannet, V., Herz, A., Lokerse, A. S., Weijers, D., Vaucheret, H., et al. (2010). miR390, *Arabidopsis* TAS3 tasiRNAs, and their AUXIN RESPONSE FACTOR targets define an autoregulatory network quantitatively regulating lateral root growth. *Plant Cell* 22, 1104–1117. doi: 10.1105/tpc.109.072553
- Martin, D. N., Proebsting, W. M., and Hedden, P. (1999). The SLENDER gene of pea encodes a gibberellin 2-oxidase. *Plant Physiol.* 121, 775–781. doi: 10.1104/pp.121.3.775
- Martínez, M., Cambra, I., González-Melendi, P., Santamaría, M. E., and Díaz, I. (2012). C1A cysteine-proteases and their inhibitors in plants. *Physiol. Plant.* 145, 85–94. doi: 10.1111/j.1399-3054.2012.01569.x
- McKim, S. M., Stenvik, G.-E., Butenko, M. A., Kristiansen, W., Cho, S. K., Hepworth, S. R., et al. (2008). The *BLADE-ON-PETIOLE* genes are essential for abscission zone formation in *Arabidopsis*. *Development* 135, 1537–1546. doi: 10.1242/dev.012807
- Medina, J., Catalá, R., and Salinas, J. (2001). Developmental and stress regulation of *RCI2A* and *RCI2B*, two cold-inducible genes of *Arabidopsis* encoding highly conserved hydrophobic proteins. *Plant Physiol.* 125, 1655–1666. doi: 10.1104/pp.125.4.1655
- Meir, S., Philosph-Hadas, S., Sundaresan, S., Selvaraj, K. S. V., Burd, S., Ophir, R., et al. (2010). Microarray analysis of the abscission-related transcriptome in the tomato flower abscission zone in response to auxin depletion. *Plant Physiol.* 154, 1929–1956. doi: 10.1104/pp.110.160697
- Mitani-Ueno, N., Yamaji, N., Zhao, F.-J., and Ma, J. F. (2011). The aromatic/arginine selectivity filter of NIP aquaporins plays a critical role in substrate selectivity for silicon, boron, and arsenic. *J. Exp. Bot.* 62, 4391–4398. doi: 10.1093/jxb/err158
- Monteoliva, M., Rizzi, Y., Cecchini, N., Hajirezaei, M.-R., and Alvarez, M. (2014). Context of action of proline dehydrogenase (ProDH) in the hypersensitive response of *Arabidopsis*. *BMC Plant Biol.* 14:21. doi: 10.1186/1471-2229-14-21
- Morillon, R., Catterou, M., Sangwan, R. S., Sangwan, B. S., and Lassalles, J.-P. (2001). Brassinolide may control aquaporin activities in *Arabidopsis thaliana*. *Planta* 212, 199–204. doi: 10.1007/s004250000379
- Nelson, D. E., Repetti, P. P., Adams, T. R., Creelman, R. A., Wu, J., Warner, D. C., et al. (2007). Plant nuclear factor Y (NF-Y) B subunits confer drought tolerance and lead to improved corn yields on water-limited acres. *Proc. Natl. Acad. Sci.* 104, 16450–16455. doi: 10.1073/pnas.0707193104
- Nishiyama, K., Guis, M., Rose, J. K. C., Kubo, Y., Bennett, K. A., Wangjin, L., et al. (2007). Ethylene regulation of fruit softening and cell wall disassembly in Charentais melon. *J. Exp. Bot.* 58, 1281–1290. doi: 10.1093/jxb/erl283
- O'Brien, M., Bertrand, C., and Matton, D. (2002). Characterization of a fertilization-induced and developmentally regulated plasma-membrane aquaporin expressed in reproductive tissues, in the wild potato *Solanum chacoense* Bitt. *Planta* 215, 485–493. doi: 10.1007/s00425-002-0770-0
- Ogawa, M., Kay, P., Wilson, S., and Swain, S. M. (2009). ARABIDOPSIS DEHISCENCE ZONE POLYGALACTURONASE1 (ADPG1), ADPG2, and QUARTET2 are polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *Plant Cell* 21, 216–233. doi: 10.1105/tpc.108.063768
- Oh, Y., Baldwin, I. T., and Galis, I. (2013). A Jasmonate ZIM-Domain protein NaJAZd regulates floral jasmonic acid levels and counteracts flower abscission in *Nicotiana attenuata* plants. *PLoS ONE* 8:e57868. doi: 10.1371/journal.pone.0057868
- Okushima, Y., Mitina, L., Quach, H. L., and Theologis, A. (2005). Auxin Response Factor2 (ARF2): a pleiotropic developmental regulator. *Plant J.* 43, 29–46. doi: 10.1111/j.1365-313X.2005.02426.x
- Ozga, J. A., and Reinecke, D. M. (2003). Hormonal interactions in fruit development. *J. Plant Growth Regul.* 22, 73–81. doi: 10.1007/s00344-003-0024-9
- Panoli, A., Martin, M. V., Alandete-Saez, M., Simon, M., Neff, C., Swarup, R., et al. (2015). Auxin import and local auxin biosynthesis are required for mitotic divisions, cell expansion and cell specification during female gametophyte development in *Arabidopsis thaliana*. *PLoS ONE* 10:e0126164. doi: 10.1371/journal.pone.0126164
- Paque, S., Mouille, G., Grandont, L., Alabadí, D., Gaertner, C., Goyallon, A., et al. (2014). AUXIN BINDING PROTEIN1 links cell wall remodeling, auxin signaling, and cell expansion in *Arabidopsis*. *Plant Cell* 26, 280–295. doi: 10.1105/tpc.113.120048
- Pekker, I., Alvarez, J. P., and Eshed, Y. (2005). Auxin response factors mediate *Arabidopsis* organ asymmetry via modulation of KANADI activity. *Plant Cell* 17, 2899–2910. doi: 10.1105/tpc.105.034876
- Petrásek, J., and Friml, J. (2009). Auxin transport routes in plant development. *Development* 136, 2675–2688. doi: 10.1242/dev.030353
- Pokotylo, I., Pejchar, P., Potocký, M., Kocourková, D., Krčková, Z., Ruelland, E., et al. (2013). The plant non-specific phospholipase C gene family. Novel competitors in lipid signalling. *Prog. Lipid Res.* 52, 62–79. doi: 10.1016/j.plipres.2012.09.001
- Porteriko, M. F., Sandaklie-Nikolova, L., Lloyd, A., Dever, C. A., Otsuga, D., and Drews, G. N. (2006). *NUCLEAR FUSION DEFECTIVE1* encodes the *Arabidopsis* RPL21M protein and is required for karyogamy during female gametophyte development and fertilization. *Plant Physiol.* 141, 957–965. doi: 10.1104/pp.106.079319
- Prusiński, J. (2007). Biological progress in lupin (*Lupinus* sp.)—present state and historical background. *Zesz. Probl. Post. Nauk Rol.* 522, 23–37.
- Qi, T., Huang, H., Song, S., and Xie, D. (2015). Regulation of jasmonate-mediated stamen development and seed production by a bHLH-MYB complex in *Arabidopsis*. *Plant Cell* 27, 1620–1633. doi: 10.1105/tpc.15.00116
- Qin, G., Gu, H., Zhao, Y., Ma, Z., Shi, G., Yang, Y., et al. (2005). An indole-3-acetic acid carboxyl methyltransferase regulates *Arabidopsis* leaf development. *Plant Cell online* 17, 2693–2704. doi: 10.1105/tpc.105.034959
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ranty, B., Aldon, D., and Galaud, J.-P. (2006). Plant calmodulins and calmodulin-related proteins: multifaceted relays to decode calcium signals. *Plant Signal. Behav.* 1, 96–104. doi: 10.4161/psb.1.3.2998
- Rebers, M., Kaneta, T., Kawaide, H., Yamaguchi, S., Yang, Y.-Y., Imai, R., et al. (1999). Regulation of gibberellin biosynthesis genes during

- flower and early fruit development of tomato. *Plant J.* 17, 241–250. doi: 10.1046/j.1365-313X.1999.00366.x
- Reidt, W., Wohlfarth, T., Ellerstrom, M., Czihal, A., Tewes, A., Ezcurra, I., et al. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the *FUS3* gene product. *Plant J.* 21, 401–408. doi: 10.1046/j.1365-313X.2000.00686.x
- Rhoades, M. W., Reinhart, B. J., Lim, L. P., Burge, C. B., Bartel, B., and Bartel, D. P. (2002). Prediction of plant microRNA targets. *Cell* 110, 513–520. doi: 10.1016/S0092-8674(02)00863-2
- Robert, S., Kleine-Vehn, J., Barbez, E., Sauer, M., Paciorek, T., Baster, P., et al. (2010). ABP1 mediates auxin inhibition of clathrin-dependent endocytosis in *Arabidopsis*. *Cell* 143, 111–121. doi: 10.1016/j.cell.2010.09.027
- Roberts, J. A., Elliott, K. A., and Gonzalez-Carranza, Z. H. (2002). Abscission, dehiscence, and other cell separation processes. *Annu. Rev. Plant Biol.* 53, 131–158. doi: 10.1146/annurev.arplant.53.092701.180236
- Roongsatham, P., Morcillo, F., Fooyontphanich, K., Jantasuriyarat, C., Tragoonrung, S., Amblard, P., et al. (2016). Cellular and pectin dynamics during abscission zone development and ripe fruit abscission of the Monocot Oil Palm. *Front. Plant Sci.* 7:540. doi: 10.3389/fpls.2016.00540
- Roongsatham, P., Morcillo, F., Jantasuriyarat, C., Pizot, M., Moussu, S., Jayaweera, D., et al. (2012). Temporal and spatial expression of polygalacturonase gene family members reveals divergent regulation during fleshy fruit ripening and abscission in the monocot species oil palm. *BMC Plant Biol.* 12:150. doi: 10.1186/1471-2229-12-150
- Rose, J. K., Lee, H. H., and Bennett, A. B. (1997). Expression of a divergent expansin gene is fruit-specific and ripening-regulated. *Proc. Natl. Acad. Sci. U.S.A.* 94, 5955–5960.
- Ross, K. J., Franz, P., Armstrong, S. J., Vizir, I., Mulligan, B., Franklin, F. C., et al. (1997). Cytological characterization of four meiotic mutants of *Arabidopsis* isolated from T-DNA-transformed lines. *Chromosome Res.* 5, 551–559.
- Sagar, M., Chervin, C., Mila, I., Hao, Y., Roustan, J.-P., Benichou, M., et al. (2013a). SLARF4, an auxin response factor involved in the control of sugar metabolism during tomato fruit development. *Plant Physiol.* 161, 1362–1374. doi: 10.1104/pp.113.213843
- Sagar, M., Chervin, C., Roustan, J.-P., Bouzayen, M., and Zouine, M. (2013b). Under-expression of the Auxin Response Factor *SL-ARF4* improves postharvest behavior of tomato fruits. *Plant Signal. Behav.* 8:e25647. doi: 10.4161/psb.25647
- Saito, T., Tuan, P. A., Katsumi-Horigane, A., Bai, S., Ito, A., Sekiyama, Y., et al. (2015). Development of flower buds in the Japanese pear (*Pyrus pyrifolia*) from late autumn to early spring. *Tree Physiol.* 35, 653–662. doi: 10.1093/treephys/tpv043
- Sakamoto, M., Munemura, I., Tomita, R., and Kobayashi, K. (2008a). Involvement of hydrogen peroxide in leaf abscission signaling, revealed by analysis with an *in vitro* abscission system in *Capsicum* plants. *Plant J.* 56, 13–27. doi: 10.1111/j.1365-313X.2008.03577.x
- Sakamoto, M., Munemura, I., Tomita, R., and Kobayashi, K. (2008b). Reactive oxygen species in leaf abscission signaling. *Plant Signal. Behav.* 3, 1014–1015. doi: 10.4161/psb.6737
- Sasse, J. M. (2003). Physiological actions of brassinosteroids: an update. *J. Plant Growth Regul.* 22, 276–288. doi: 10.1007/s00344-003-0062-3
- Sawicki, M., Ait Barka, E., Clement, C., Vaillant-Gaveau, N., and Jacquard, C. (2015). Cross-talk between environmental stresses and plant metabolism during reproductive organ abscission. *J. Exp. Bot.* 66, 1707–1719. doi: 10.1093/jxb/eru533
- Schott, G., Mari-Ordóñez, A., Himber, C., Alioua, A., Voinnet, O., and Dunoyer, P. (2012). Differential effects of viral silencing suppressors on siRNA and miRNA loading support the existence of two distinct cellular pools of ARGONAUTE1. *Embo J.* 31, 2553–2565. doi: 10.1038/emboj.2012.92
- Schruff, M. C., Spielman, M., Tiwari, S., Adams, S., Fenby, N., and Scott, R. J. (2005). The *AUXIN RESPONSE FACTOR2* gene of *Arabidopsis* links auxin signalling, cell division, and the size of seeds and other organs. *Development* 133, 251–261. doi: 10.1242/dev.02194
- Schumacher, K., Vafeados, D., McCarthy, M., Sze, H., Wilkins, T., and Chory, J. (1999). The *Arabidopsis det3* mutant reveals a central role for the vacuolar H⁺-ATPase in plant growth and development. *Genes Dev.* 13, 3259–3270.
- Serrani, J. C., Ruiz-Rivero, O., Fos, M., and García-Martínez, J. L. (2008). Auxin-induced fruit-set in tomato is mediated in part by gibberellins. *Plant J.* 56, 922–934. doi: 10.1111/j.1365-313X.2008.03654.x
- Serrani, J. C., Sanjuán, R., Ruiz-Rivero, O., Fos, M., and García-Martínez, J. L. (2007). Gibberellin regulation of fruit set and growth in tomato. *Plant Physiol.* 145, 246–257. doi: 10.1104/pp.107.098335
- Shimada, Y., Fujioka, S., Miyauchi, N., Kushiro, M., Takatsuto, S., Nomura, T., et al. (2001). Brassinosteroid-6-oxidases from *Arabidopsis* and tomato catalyze multiple C-6 oxidations in brassinosteroid biosynthesis. *Plant Physiol.* 126, 770–779. doi: 10.1104/pp.126.2.770
- Singh, A. P., Tripathi, S. K., Nath, P., and Sane, A. P. (2011). Petal abscission in rose is associated with the differential expression of two ethylene-responsive xyloglucan endotransglucosylase/hydrolase genes, *RbXTH1* and *RbXTH2*. *J. Exp. Bot.* 62, 5091–5103. doi: 10.1093/jxb/err209
- Singh, K., Talla, A., and Qiu, W. (2012). Small RNA profiling of virus-infected grapevines: evidences for virus infection-associated and variety-specific miRNAs. *Funct. Integr. Genomics* 12, 659–669. doi: 10.1007/s10142-012-0292-1
- Sipes, D. L., and Einset, J. W. (1983). Cytokinin stimulation of abscission in lemon pistil explants. *J. Plant Growth Regul.* 2, 73–80. doi: 10.1007/BF02042235
- Sitrit, Y., and Bennett, A. B. (1998). Regulation of tomato fruit polygalacturonase mRNA accumulation by ethylene: a Re-examination. *Plant Physiol.* 116, 1145–1150.
- Skoog, F., Strong, F. M., and Miller, C. O. (1965). Cytokinins. *Science* 148, 532–533. doi: 10.1126/science.148.3669.532-a
- Song, H., Zhao, X., Hu, W., Wang, X., Shen, T., and Yang, L. (2016). Comparative transcriptional analysis of loquat fruit identifies major signal networks involved in fruit development and ripening process. *Int. J. Mol. Sci.* 17, 1837. doi: 10.3390/ijms17111837
- Stenvik, G.-E., Tandstad, N. M., Guo, Y., Shi, C.-L., Kristiansen, W., Holmgren, A., et al. (2008). The EPIP peptide of INFLORESCENCE DEFICIENT IN ABCISSION is sufficient to induce abscission in *Arabidopsis* through the receptor-like kinases HAESA and HAESA-LIKE2. *Plant Cell* 20, 1805–1817. doi: 10.1105/tpc.108.059139
- Sundaresan, S., Philosoph-Hadas, S., Riov, J., Mugasimangalam, R., Kuravadi, N. A., Kochanek, B., et al. (2016). *De novo* transcriptome sequencing and development of abscission zone-specific microarray as a new molecular tool for analysis of Tomato organ abscission. *Front. Plant Sci.* 6:1258. doi: 10.3389/fpls.2015.01258
- Sunkar, R., and Jagadeeswaran, G. (2008). *In silico* identification of conserved microRNAs in large number of diverse plant species. *BMC Plant Biol.* 8:37. doi: 10.1186/1471-2229-8-37
- Tadeo, F. R., and Primo-Millo, E. (1990). Peroxidase activity changes and lignin deposition during the senescence process in citrus stigmas and styles. *Plant Sci.* 68, 47–56. doi: 10.1016/0168-9452(90)90151-D
- Takano, J., Wada, M., Ludewig, U., Schaaf, G., von Wirén, N., and Fujiwara, T. (2006). The *Arabidopsis* major intrinsic protein NIP5;1 is essential for efficient boron uptake and plant development under boron limitation. *Plant Cell Online* 18, 1498–1509. doi: 10.1105/tpc.106.041640
- Taylor, J. E., and Whitelaw, C. A. (2001). Signals in abscission. *New Phytol.* 151, 323–340. doi: 10.1046/j.0028-646x.2001.00194.x
- Tran, L.-S. P., Nakashima, K., Sakuma, Y., Simpson, S. D., Fujita, Y., Maruyama, K., et al. (2004). Isolation and functional analysis of *Arabidopsis* stress-inducible NAC transcription factors that bind to a drought-responsive *cis*-element in the early response to dehydration stress 1 promoter. *Plant Cell* 16, 2481–2498. doi: 10.1105/tpc.104.022699
- Tsuchiya, M., Satoh, S., and Iwai, H. (2015). Distribution of XTH, expansin, and secondary-wall-related CesA in floral and fruit abscission zones during fruit development in tomato (*Solanum lycopersicum*). *Front. Plant Sci.* 6:323. doi: 10.3389/fpls.2015.00323
- Tucker, M. L., Burke, A., Murphy, C. A., Thai, V. K., and Ehrenfried, M. L. (2007). Gene expression profiles for cell wall-modifying proteins associated with soybean cyst nematode infection, petiole abscission, root tips, flowers, apical buds, and leaves. *J. Exp. Bot.* 58, 3395–3406. doi: 10.1093/jxb/erm188
- Unamba, C. I. N., Nag, A., and Sharma, R. K. (2015). Next generation sequencing technologies: the doorway to the unexplored genomics of non-model plants. *Front. Plant Sci.* 6:1074. doi: 10.3389/fpls.2015.01074
- UniProt Consortium (2015). UniProt: a hub for protein information. *Nucleic Acids Res.* 43, D204–D212. doi: 10.1093/nar/gku989
- Vaddepalli, P., Herrmann, A., Fulton, L., Oelschner, M., Hillmer, S., Stratil, T. F., et al. (2014). The C2-domain protein QUIRKY and the receptor-like kinase

- STRUBBELIG localize to plasmodesmata and mediate tissue morphogenesis in *Arabidopsis thaliana*. *Development* 141, 4139–4148. doi: 10.1242/dev.113878
- van Doorn, W. G., Beers, E. P., Dangel, J. L., Franklin-Tong, V. E., Gallois, P., Hara-Nishimura, I., et al. (2011). Morphological classification of plant cell deaths. *Cell Death Differ.* 18, 1241–1246. doi: 10.1038/cdd.2011.36
- Van Steveninck, R. F. (1958). Factors affecting the abscission of reproductive organs in yellow lupins (*Lupinus luteus* L.) II. The effects of growth substances, defoliation, and removal of lateral growth. *J. Exp. Bot.* 37, 373–383.
- Van Steveninck, R. F. (1959). Abscission-accelerators in lupins (*Lupinus luteus* L.). *Nature* 183, 1246–1248.
- Vanneste, S., and Friml, J. (2009). Auxin: a trigger for change in plant development. *Cell* 136, 1005–1016. doi: 10.1016/j.cell.2009.03.001
- Vert, G., Nemhauser, J. L., Geldner, N., Hong, F., and Chory, J. (2005). Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* 21, 177–201. doi: 10.1146/annurev.cellbio.21.090704.151241
- Vetten, N. C., and Huber, D. J. (1990). Cell wall changes during the expansion and senescence of carnation (*Dianthus caryophyllus*) petals. *Physiol. Plant.* 78, 447–454. doi: 10.1111/j.1399-3054.1990.tb09062.x
- Voinnet, O. (2009). Origin, biogenesis, and activity of plant microRNAs. *Cell* 136, 669–687. doi: 10.1016/j.cell.2009.01.046
- Vriezen, W. H., Feron, R., Maretto, F., Keijman, J., and Mariani, C. (2007). Changes in tomato ovary transcriptome demonstrate complex hormonal regulation of fruit set. *New Phytol.* 177, 60–76. doi: 10.1111/j.1469-8137.2007.02254.x
- Wallace, I. S., and Roberts, D. M. (2005). Distinct transport selectivity of two structural subclasses of the nodulin-like intrinsic protein family of plant aquaglyceroporin channels. *Biochemistry* 44, 16826–16834. doi: 10.1021/bi0511888
- Wang, F., and Perry, S. E. (2013). Identification of direct targets of FUSCA3, a key regulator of *Arabidopsis* seed development. *Plant Physiol.* 161, 1251–1264. doi: 10.1104/pp.112.212282
- Wang, M., Zhang, X., and Liu, J. H. (2015). Deep sequencing-based characterization of transcriptome of trifoliolate orange (*Poncirus trifoliata* (L.) Raf.) in response to cold stress. *BMC Genomics* 16:555. doi: 10.1186/s12864-015-1629-7
- Wang, Y., Zhu, S., Liu, S., Jiang, L., Chen, L., Ren, Y., et al. (2009). The vacuolar processing enzyme OsVPE1 is required for efficient glutelin processing in rice. *Plant J.* 58, 606–617. doi: 10.1111/j.1365-313X.2009.03801.x
- Warpeha, K. M., Upadhyay, S., Yeh, J., Adamiak, J., Hawkins, S. I., Lapik, Y. R., et al. (2007). The GCR1, GPA1, PRN1, NF-Y signal chain mediates both blue light and abscisic acid responses in *Arabidopsis*. *Plant Physiol.* 143, 1590–1600. doi: 10.1104/pp.106.089904
- Wasternack, C. (2007). “Oxylipins: biosynthesis, signal transduction and action,” in *Annual Plant Reviews Vol. 24, Plant Hormone Signaling*, eds P. Hedden and S. G. Thomas (Oxford: Blackwell Publishing Ltd.), 185–228.
- Wenkel, S., Turck, F., Singer, K., Gissot, L., Le Gourrierc, J., Samach, A., et al. (2006). CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of *Arabidopsis*. *Plant Cell* 18, 2971–2984. doi: 10.1105/tpc.106.043299
- Werner, T., Motyka, V., Laucou, V., Smets, R., Van Onckelen, H., and Schmülling, T. (2003). Cytokinin-deficient transgenic *Arabidopsis* plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. *Plant Cell* 15, 2532–2550. doi: 10.1105/tpc.014928
- White, P. J., and Broadley, M. R. (2003). Calcium in Plants. *Ann. Bot.* 92, 487–511. doi: 10.1093/aob/mcg164
- Wilmowicz, E., Frankowski, K., Kučko, A., Świdziński, M., de Dios Alché, J., Nowakowska, A., et al. (2016). The influence of abscisic acid on the ethylene biosynthesis pathway in the functioning of the flower abscission zone in *Lupinus luteus*. *J. Plant Physiol.* 206, 49–58. doi: 10.1016/j.jplph.2016.08.018
- Won, C., Shen, X., Mashiguchi, K., Zheng, Z., Dai, X., Cheng, Y., et al. (2011). Conversion of tryptophan to indole-3-acetic acid by tryptophan aminotransferases of *Arabidopsis* and YUCCAs in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 18518–18523. doi: 10.1073/pnas.1108436108
- Yamada, Y., and Umehara, M. (2015). Possible roles of strigolactones during leaf senescence. *Plants (Basel)* 4, 664–677. doi: 10.3390/plants4030664
- Yamaguchi, N., Suzuki, M., Fukaki, H., Morita-Terao, M., Tasaka, M., and Komeda, Y. (2007). CRM1/BIG-mediated auxin action regulates *Arabidopsis* inflorescence development. *Plant Cell Physiol.* 48, 1275–1290. doi: 10.1093/pcp/pcm094
- Yang, H., Tao, Y., Zheng, Z., Zhang, Q., Zhou, G., Sweetingham, M. W., et al. (2013). Draft genome sequence, and a sequence-defined genetic linkage map of the legume crop species *Lupinus angustifolius* L. *PLoS ONE* 8:e64799. doi: 10.1371/journal.pone.0064799
- Yin, P., Fan, H., Hao, Q., Yuan, X., Wu, D., Pang, Y., et al. (2009). Structural insights into the mechanism of abscisic acid signaling by PYL proteins. *Nat. Struct. Mol. Biol.* 16, 1230–1236. doi: 10.1038/nsmb.1730
- Zhang, K., and Gan, S.-S. (2012). An abscisic acid-AtNAP transcription factor-SAG113 protein phosphatase 2C regulatory chain for controlling dehydration in senescing *Arabidopsis* leaves. *Plant Physiol.* 158, 961–969. doi: 10.1104/pp.111.190876
- Zhao, M., Ding, H., Zhu, J.-K., Zhang, F., and Li, W.-X. (2011). Involvement of miR169 in the nitrogen-starvation responses in *Arabidopsis*. *New Phytol.* 190, 906–915. doi: 10.1111/j.1469-8137.2011.03647.x
- Zhao, Z., Zhang, Y., Liu, X., Zhang, X., Liu, S., Yu, X., et al. (2013). A role for a dioxygenase in auxin metabolism and reproductive development in rice. *Dev. Cell* 27, 113–122. doi: 10.1016/j.devcel.2013.09.005
- Zhou, C., Cai, Z., Guo, Y., and Gan, S. (2009). An *Arabidopsis* mitogen-activated protein kinase cascade, MKK9-MPK6, plays a role in leaf senescence. *Plant Physiol.* 150, 167–177. doi: 10.1104/pp.108.133439
- Zhou, L., Liu, Y., Liu, Z., Kong, D., Duan, M., and Luo, L. (2010). Genome-wide identification and analysis of drought-responsive microRNAs in *Oryza sativa*. *J. Exp. Bot.* 61, 4157–4168. doi: 10.1093/jxb/erq237
- Zhou, Y., Yang, P., Cui, F., Luo, X., and Xie, J. (2016). Transcriptome analysis of salt stress responsiveness in the seedlings of dongxiang wild rice (*Oryza rufipogon* Griff.). *PLoS ONE* 11:e0146242. doi: 10.1371/journal.pone.0146242

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Glazinska, Wojciechowski, Kulasek, Glinkowski, Marciniak, Klajn, Kesy and Kopcewicz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of *ZOUPI* Orthologs in Soybean Potentially Involved in Endosperm Breakdown and Embryogenic Development

Yaohua Zhang, Xin Li, Suxin Yang* and Xianzhong Feng*

Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Spanish National Research Council,
Spain

Reviewed by:

Gwyneth Ingram,
École Normale Supérieure de Lyon,
France
Francisco A.P. Campos,
Federal University of Ceará, Brazil

*Correspondence:

Suxin Yang
yangsuxin@iga.ac.cn
Xianzhong Feng
fengxianzhong@iga.ac.cn

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 11 October 2016

Accepted: 23 January 2017

Published: 08 February 2017

Citation:

Zhang Y, Li X, Yang S and Feng X
(2017) Identification of *ZOUPI*
Orthologs in Soybean Potentially
Involved in Endosperm Breakdown
and Embryogenic Development.
Front. Plant Sci. 8:139.
doi: 10.3389/fpls.2017.00139

Soybean (*Glycine max* Merr.) is the world's most widely grown legume and provides an important source of protein and oil. Improvement of seed quality requires deep insights into the genetic regulation of seed development. The endosperm serves as a temporary source of nutrients that are transported from maternal to filial tissues, and it also generates signals for proper embryo formation. Endosperm cell death is associated with the processes of nutrient transfer and embryo expansion. The bHLH domain transcription factor *AtZHOUPI* (*AtZOU*) plays a key role in both the lysis of the transient endosperm and the formation of embryo cuticle in *Arabidopsis thaliana*. There are two copies of soybean *GmZOU* (*GmZOU-1* and *GmZOU-2*), which fall into the same phylogenetic clade as *AtZOU*. These two copies share the same transcription orientation and are the result of tandem duplication. The expression of *GmZOUS* is limited to the endosperm, where it peaks during the heart embryo stage. When the exogenous *GmZOU-1* and *GmZOU-2* were expressed in the *zou-4* mutant of *Arabidopsis*, only *GmZOU-1* partially complemented the *zou* mutant phenotype, as indicated by endosperm breakdown and embryo cuticle formation in the transgenic lines. This research confirmed that the *GmZOU-1* is a *ZOU* ortholog that may be responsible for endosperm breakdown and embryo cuticle formation in soybean.

Keywords: seed development, endosperm breakdown, transcription factor, *GmZOU*, soybean

INTRODUCTION

Soybean is a major global economic crop and a source of carbohydrates, protein, oil, and other nutrients for humans and animals (Hou et al., 2009). Improvements of the production and seed quality of soybean require a thorough understanding of the underlying processes governing seed growth and development. However, most of the research performed on soybean have focused on various biochemical pathways for producing seed storage products based on known enzymatic steps, and less attention has been given on the relationship between seed development and storage product metabolic programs, e.g., coordination of different seed compartments pertaining to resource channeling and storage product accumulation, and as such, the regulatory and dynamic aspects of the network are still poorly understood.

In dicotyledonous species, the endosperm plays important roles in metabolite production, transport, and accumulation in the embryo (Melkus et al., 2009). The endosperm in Brassicaceae

is believed to participate in the flux control of nutrients delivered by the vascular tissues of the parental plants to the embryo (Abbadì and Leckband, 2011). The soluble metabolites are temporarily stored in the endosperm vacuole with the concentration developmentally controlled, and influence metabolite accumulation in embryo (Melkus et al., 2009). The endosperm is also involved in lipid, sugar, amino acid, and organic acid metabolism (Schwender and Ohlrogge, 2002; Hill et al., 2003), and in mineral acquisition and storage (Otegui et al., 2002). In addition to nutrient transport and metabolism, the endosperm also plays a role in controlling the sizes of the embryo and seed, although the precise mechanism is still unknown (Luo et al., 2005).

Endosperm development in most angiosperms is of the nuclear type and is characterized by four phases: syncytial, cellularization, differentiation, and death (Olsen, 2004). After cellularization, the embryo surrounding region (ESR) of the endosperm cell starts to breakdown, thereby freeing the nutrients that fuel the embryo and in turn create space for embryo expansion (Ingram, 2010). In dicotyledons such as *Arabidopsis* and soybean, the endosperm undergoes programmed cell death with complete autolysis, leaving a single cell layer of endosperm tissue surrounding the dormant embryo (Ingram, 2010). The endosperm death stage, which represents the beginning of the maturation stage, refers to a process involving nutrient transfer and storage in the embryo, as well as the generation of developmental signals that are transmitted to the young embryo (Berger, 2003; Sharma et al., 2003; Berger et al., 2006; Baud et al., 2008). During this stage, the sucrose transporter *AtSUC5* is induced in the ESR region and plays an important yet transient role in the transport of nutrients to the embryo (Baud et al., 2005). In pea and *Vicia* spp. seeds, a change from a predominantly hexose to sucrose content in the endosperm induces a change of gene expression in favor of storage product accumulation (Hill et al., 2003). *SBT1.1*, a subtilase gene, is specifically expressed in the endosperm of *Medicago truncatula* and *Pisum sativum* seeds to control embryo growth (D'Erfurth et al., 2012). The endosperm degradation stage is a transition process that influences seed development and metabolite accumulation.

The *ZOUPI* (*ZOU*) gene is a unique and highly conserved bHLH transcription factor that is exclusively expressed in the ESR and controls both endosperm breakdown and embryonic cuticle formation during *Arabidopsis* seed development (Kondou et al., 2008; Yang et al., 2008). In *zou/rge1* mutants, ESR cell death is compromised, and the endosperm persists in the mature seeds with heart shaped embryo. In addition the defect in the embryonic cuticle of *zou* mutants lead to splits in the cotyledon epidermis after germination. The *ZOU* gene is also specifically expressed in the maize endosperm, influencing ESR breakdown and embryonic and seed development (Grimault et al., 2015). The conserved function of the *ZOU* gene in both *Arabidopsis* and maize indicates that it plays an important role in the communication between endosperm breakdown and embryo development.

In order to investigate the regulating pathway of endosperm breakdown and the communication between endosperm and embryo in soybean, we identified a *ZOU* ortholog gene, which

complements the *Arabidopsis zou* mutant phenotype allowing the recovery of endosperm breakdown and embryo cuticle formation.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Soybean plants [*G. max* (L.) Merrill] (Williams 82) were grown under natural conditions in the field at the Agricultural Experiment Station of Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China. The *Arabidopsis zou* mutant *zou-4* (Col-0 background) allele was obtained from the Gabi-Kat collection of T-DNA inserts (Rosso et al., 2003) and corresponded to GABI_584D09. *Arabidopsis* seeds were surface-sterilized, vernalized at 4°C for 3 days, and then allowed to germinate and grow on plant growth medium in the growth chamber maintained under a 16-h light/8-h dark (22–18°C) photoperiod. Soil-grown plants were also maintained under the same conditions.

GmZOU Cloning by RT-PCR

Total RNA was isolated from soybean developing seeds under normal growth conditions using a Plant Total RNA Extraction Kit (Bioteke Corporation) and first strand cDNA was prepared from 0.5 µg of total RNA primed with an oligo-dT primer using an AMV reverse transcription system (Takara) according to the manufacturer's instructions. The genes were cloned using PrimeSTAR HS DNA Polymerase (Takara). Two PCR primers, OL0085 and OL0086, were designed to amplify the *GmZOU-1* gene. A 937-bp PCR fragment was amplified, and the PCR reactions were performed as follows: 98°C for 3 min and then 30 cycles of 98°C for 10 s, 55°C for 15 s, and 72°C for 1 min. Primers OL0087 and OL0088 were designed to amplify the *GmZOU-2* gene. A 830-bp PCR fragment was amplified and the PCR reactions were performed as follows: 98°C for 3 min and then 30 cycles of 98°C for 10 s, 55°C for 15 s, and 72°C for 50 s. 3' A overhangs were attached to the PCR fragments by using EasyTaq (TransGenBiotech) before their cloning into a pMD18-T vector (Takara).

Phylogenetic Analysis

The *ZOU* protein sequences from different species with similarities to AtZOU were mined from Phytozome V11¹ and GenBank (National Institutes of Health genetic sequence database). The amino acid sequences were aligned using CLUSTALW (Thompson et al., 1994). We limited the boundaries of the sequences. In total, 21 sequences (Supplementary Dataset 1) were analyzed. Phylogenetic reconstruction of the *ZOU* genes was performed using the neighbor-joining (NJ) method (Saitou and Nei, 1987). The bootstrap consensus tree inferred from 1,000 replicates was used to represent the evolutionary history of the analyzed taxa (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates were collapsed. The percentage of replicate trees in which the

¹<https://phytozome.jgi.doe.gov>

associated taxa were clustered together in the bootstrap test (1,000 replicates) was shown next to the branches (Felsenstein, 1985). The tree was drawn to scale, with branch lengths in the same units as those of the evolutionary distances that were used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. All positions containing alignment gaps and missing data were eliminated only in pairwise sequence comparisons (pairwise deletion option). A total of 451 positions were included in the final dataset. Phylogenetic analyses were conducted using MEGA4 (Tamura et al., 2007).

Quantitative Gene Expression Analysis

In soybean, total RNA was extracted from different tissues, including root, stem, leaf, inflorescence, unopened flower, opened flower, and seeds at 5 DAP (globular stage), 9 DAP (early heart stage), 12 DAP (late heart stage), and 16 DAP (cotyledon stage; Jones and Vodkin, 2013). To obtain the endosperm and embryo separately from soybean seeds, the seeds were dissected transversely (slightly out of the middle toward the micropylar end) using forceps and a microtome blade (Leica). The embryo was picked out by forceps and the endosperm was separated with integument using a scalpel knife under the stereo microscope (Olympus SZX7), and material was flash frozen in liquid nitrogen. In *Arabidopsis*, total RNA was extracted from siliques with seeds developing to the heart stage (about 5 DAP). The total RNA was treated with DNase I (Takara) according to the manufacturer's instructions. RNA concentrations were determined by using a Nanodrop Spectrophotometer (NanoDrop Technologies). First-strand cDNA was prepared from 0.5 µg of total RNA primed with oligo-dT primer using an AMV reverse transcription system (Takara). PCR reactions were prepared using SYBR Premix EX Taq (Takara) and each 25 µl reaction was triplicated (technical replicates) and for each experiment three biological replicates (i.e., independent plant samples or different plants in the same transgenic line) were made. According to the manufacturer's protocol, the following program was used: 10 min at 95°C, followed by 40 cycles of 95°C for 10 s and 60°C for 1 min. The products were quantified using a Bio-Rad DNA Engine Opticon 2 real-time PCR machine and an associated software to assay SYBR green fluorescence. Expression levels were calculated using *ACTIN11* (Soybean), which is a commonly used reference gene with stable expression in different tissues (Hu et al., 2009), and *EIF4A1* (*Arabidopsis*) in developing seed as described (Yang et al., 2008; Xing et al., 2013; Grimault et al., 2015). The primers used are listed in Supplementary Table S1.

Vector Construction and Plant Transformation

To construct the *pAtZOU::GmZOU-1* and *pAtZOU::GmZOU-2* expression vectors, cloned *GmZOU-1* and *GmZOU-2* CDS in pMD18-T vectors in the correct orientation were digested with *Sall* and *SmaI* and cloned into the vector pRT101 that was digested by *XhoI* and *HincII*, thereby constructing the vectors ZYH051 and ZYH052. ZYH051 and ZYH052 were then digested

with *PstI*, and the gene fragments were separately cloned into a binary vector pCambia1300 to construct the vectors, ZYH053 and ZYH054. A 1.6-kb region upstream from the *AtZOU* start codon was amplified by PCR (as described earlier) using primers ZOUSALIF and ZOUSALIR and subcloned into vector pSC-B (Stratagene). The vector containing *AtZOU* promoter was digested with *Sall*, and the promoter fragment was then cloned into the binary vectors ZYH053 and ZYH054, thereby constructing the *pAtZOU::GmZOU-1* and *pAtZOU::GmZOU-2* expression vectors ZYH055 and ZYH056. All recombinant plasmids identified from individual *Escherichia coli* colonies were verified by sequencing, the expression vectors described above were introduced into *Agrobacterium tumefaciens* GV3101 independently for plant transformation. Plant transformations were conducted according to Clough and Bent (1998). Plants were dipped twice in a solution of 10% sucrose, 0.5 × MS salts, and 0.05% silwet L-77 solution containing GV3101 *Agrobacterium* cells that were grown for 48 h.

Selection of Transformed Plants

Seeds from the transformed plants were screened on MS agar (Sigma) containing 2% sucrose, 20 mg L⁻¹ hygromycin (Roche) and 100 mg L⁻¹ cefotaxime (Sangong). Resistant plants were transferred to pots with soil and perlite (3:1 v/v) and grown to maturity in a growth chamber. Hygromycin-resistant plants were screened for the presence of the transgene by PCR using soybean *GmZOU*s-specific primers (above). Seeds were harvested and sown on MS agar containing 2% sucrose, 20 mg L⁻¹ hygromycin to obtain homozygous plants. Homozygous plants were transferred to soil and were grown for seed production. Seeds from these homozygous plants were used in the subsequent experiments.

Toluidine Blue Staining

The Toluidine Blue (TB) test was carried out following published procedures (Tanaka et al., 2004; Xing et al., 2013). Seeds were spread uniformly on 15-cm plates containing 1 × MS Basal Salts (Sigma), 0.3% sucrose, and 0.4% Phytigel (Sigma; pH 5.8). Stratification was conducted at 4°C for 3 days before transferring plates to a growth chamber for 7 days. Lids were removed and plates were immediately flooded with the staining solution [0.05% (w/v) TB + 0.4% (v/v) Tween-20] for 2 min. The staining solution was poured off, and the plates were immediately rinsed gently by flooding under a running tap until the water cleared (1–2 min). Seedlings were photographed, or harvested for TB quantification. To harvest, seedlings were removed individually from plates and both roots and any adhering seed coats (both of which stain darkly with TB) were completely removed before plunging the hypocotyl and cotyledons into 1 ml of 80% ethanol. Seedlings were incubated with continuous shaking for 2 h, until all blue color and chlorophyll had been removed from cotyledons. The resulting liquid was analyzed using a spectrophotometer.

Seed Clearing

To visualize and stage the developing seeds, siliques were opened with needles. The seeds were then cleared overnight in modified Hoyer's solution (chloral hydrate: water: glycerol in proportions

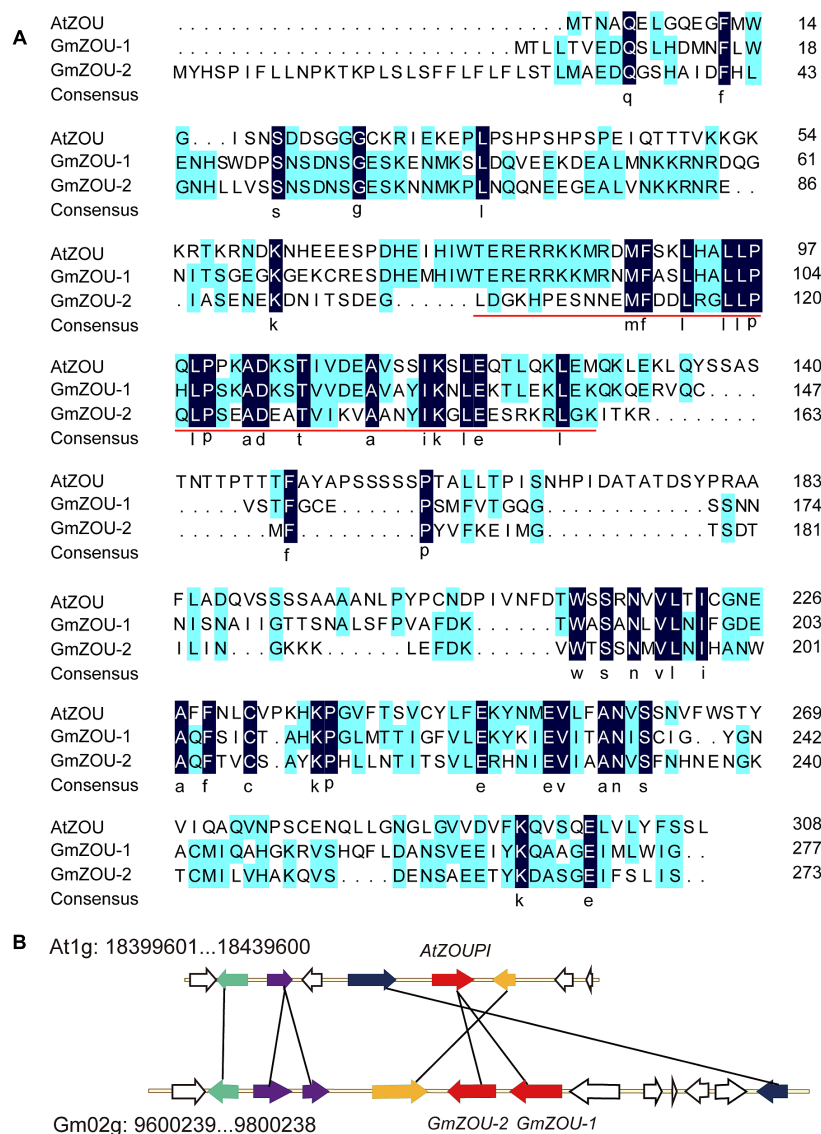


FIGURE 1 | Identification of ZOU homologs in soybean. (A) Alignment of amino acid sequences of the AtZOU gene with GmZOU-1 (Glyma.02G103200) and GmZOU-2 (Glyma.02G103100). Identical amino acids are indicated by white letters against a black background and the conserved bHLH domain (pfam00010) was remarked by red line. **(B)** A syntenic block of soybean and *Arabidopsis thaliana* sequence assemblies surrounding the *Arabidopsis* ZOU gene. Solid black lines connect homolog gene pairs and are indicate by the same color. White arrows represent genes with no syntentic homolog(s) in the particular genomic region. The ZOUPI tandem duplicates were detected in soybean (GmZOU-1 and GmZOU-2).

8 g : 2 mL : 1 mL) and visualized under DIC optics using a Nikon Eclipse E600 microscope.

RESULTS

GmZOU Genes Cloned from Developing Seeds of Soybean

To identify the ZOU homologs in soybean, the TBLASTN program was used to query the deduced amino acid sequences of gene models from the soybean genomic sequence database (Phytozome) with that of AtZOU. The two highest scores

were obtained for the genes models *Glyma.02G103200* and *Glyma.02G103100*, and the corresponding genes were designated as *GmZOU-1* and *GmZOU-2*, respectively. Alignment of the amino acid sequences of the *GmZOU-1* and *GmZOU-2* with that of AtZOU revealed the presence of a basic helix-loop-helix (bHLH) DNA-binding domain (pfam00010) and a conserved C-terminal domain (Figure 1A). Phylogenetic analysis of the all 480 bHLH family proteins in soybean and the AtZOU protein showed that both *GmZOU*s were in the same phylogenetic clade with AtZOU (Supplementary Figure S1).

The two copies of *GmZOU*s identified in soybean were located on Gm02: 9,795,987–9,799,196 (Accession Number

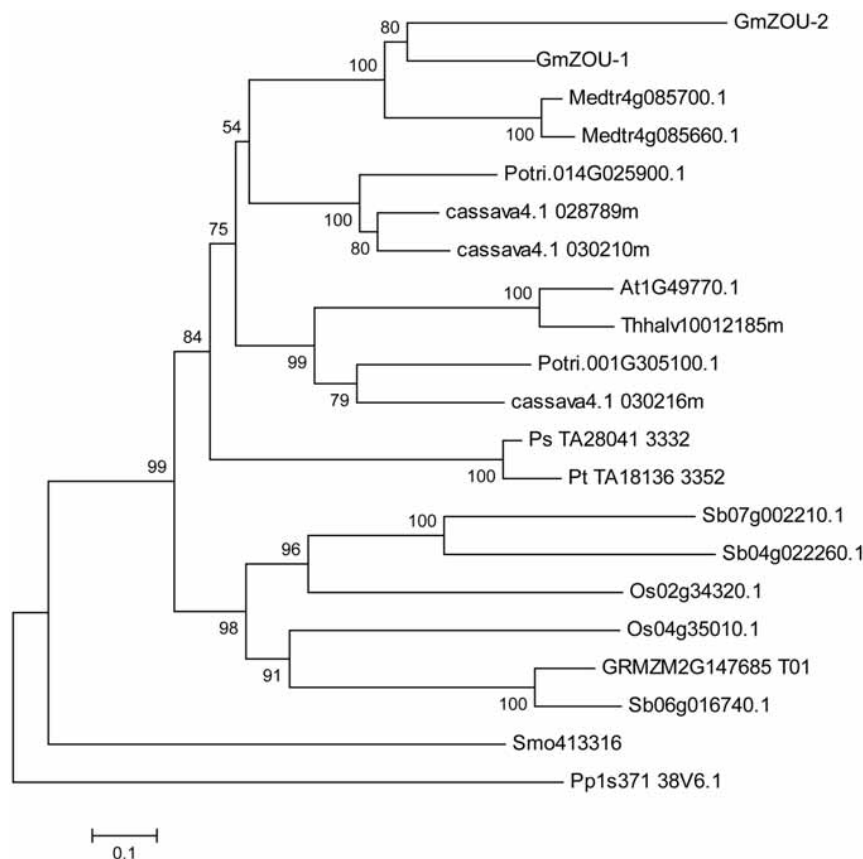


FIGURE 2 | Phylogenetic analysis of ZOU. A Neighbor-joining tree of ZOU homologs from *Arabidopsis* (*A. thaliana*, At), salt cress (*T. halophila*, Thhalv), soybean (*G. max*, Glyma), lucerne (*M. truncatula*, Medtr), poplar (*P. trichocarpa*, Potri), cassava (*M. esculenta*, cassava), rice (*O. sativa*, Os), corn (*Z. mays*, GRMZM), sorghum (*S. bicolor*, Sb), pine (*P. taeda*, Pt), spruce (*P. sitchensis*, Ps), selaginella (*S. moellendorffii*, Smo), and moss (*P. patens*, Pp). Bootstrapping values are indicated where the local bootstrap probability of each branch is > 50.

Glyma.02G103200) and Gm02: 9,784,634–9,786,935 (Accession Number Glyma.02G103100), which is adjacent to the chromosome with the same transcription orientation. The chromosome fragment in soybean chromosome 2 containing 13 genes around the *GmZOU*s was aligned with the chromosome fragment containing *AtZOU* from the *Arabidopsis* genome. There were five genes in this selected *Arabidopsis* chromosome region that corresponded to seven homologs in this soybean chromosomal segment that contained two genes that were tandemly duplicated. In addition, these homologs showed synteny, as indicated by the same order and orientation (Figure 1B). The two chromosomal regions showed similar evolutionary origin, and the two copies of *GmZOU*s were the result of tandem gene duplication.

Phylogenetic Analysis of *GmZOU*s

To identify the phylogenetic relationships and functional conservation of different species during evolution, the homolog of the *ZOU* gene in other species, including *A. thaliana*, salt cress (*Thellungiella halophila*), soybean (*Glycine max*), lucerne (*M. truncatula*), poplar (*Populus trichocarpa*), cassava (*Manihot esculenta*), rice (*Oryza sativa*), corn (*Zea mays*),

sorghum (*Sorghum bicolor*), pine (*Pinus taeda*), spruce (*Picea sitchensis*), selaginella (*Selaginella moellendorffii*), and moss (*Physcomitrella patens*), were downloaded and filtered. Only full-length sequences and those with at least 30% similarity with the *AtZOU* gene were included in the analysis. A total of 21 sequences were selected and used in the subsequent NJ phylogenetic analysis. A strict consensus tree was also generated (Figure 2).

The phylogenetic tree showed that *ZOU* is widely conserved in plants. Within angiosperms, *ZOU* was found in both plants that develop persistent endosperm and plants that do not as well as in seeded plants lacking endosperm (e.g., *P. sitchensis*, a gymnosperm). It was also found in more basal vascular plant groups, such as the moss (e.g., *P. patens*) and ferns (e.g., *S. moellendorffii*), which lack seeds altogether. Besides that the *ZOU* genes in gymnosperms such as *P. sitchensis* and *P. taeda* had a closer relationship with dicotyledonous species compared to monocotyledonous species. This may be related to the specific function of the *ZOU* gene, which regulated the degeneration of maternal material during gametogenesis. In monocotyledonous species, the gametogenesis evolved in a specific manner wherein the maternally derived nutrients are stored in the endosperm

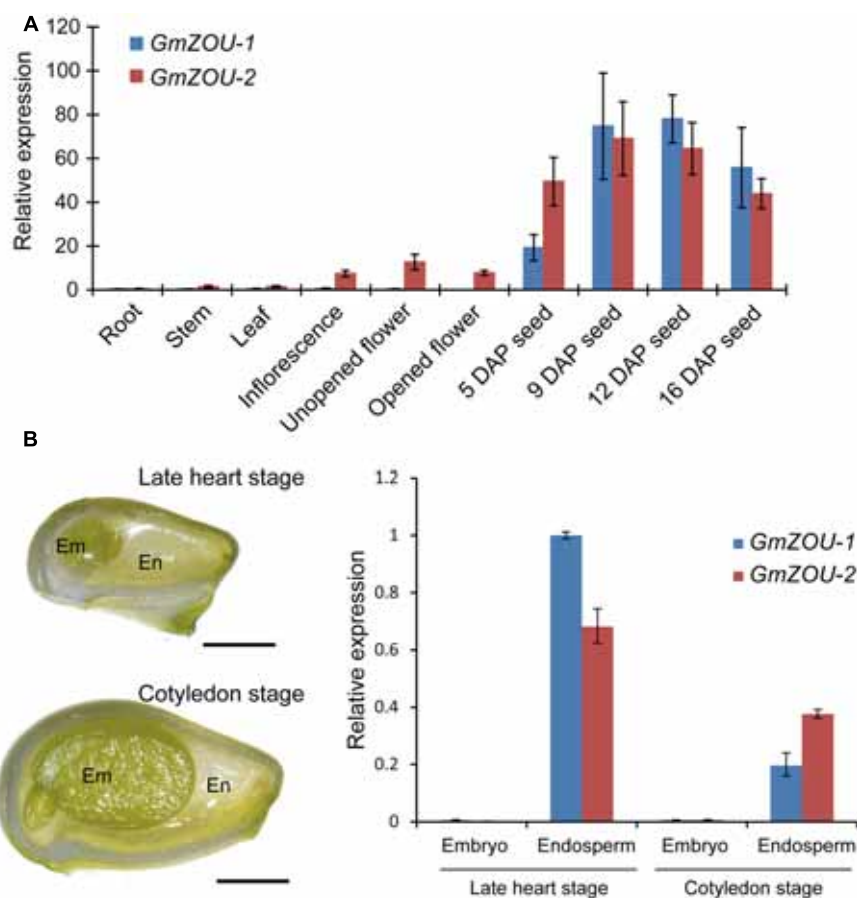


FIGURE 3 | Expression analyses of *GmZOU-1* and *GmZOU-2* in soybean. (A) The expression of *GmZOU*s in different tissues, including the root, stem, leaf, inflorescence, unopened flower, opened flower, and seeds at different DAP, as detected by qPCR analysis. **(B)** qPCR analysis of *GmZOU*s expression on dissected seed compartments. The embryo and endosperm were separately isolated from developing seeds at late heart stage or cotyledon stage. For each stage at least 20 seeds were selected with three biological replicates. Scale bars: 1 mm.

until seed germination. Although ZOU in monocotyledonous species, like *AtZOU* is involved in the local lysis of endosperm to allow embryo growth, the bulk of persistent endosperm implies some functional alterations. Recently the potential functional specificity of *ZmZOU* unique to monocotyledon had also been shown in maize (Grimault et al., 2015).

Expression Profiles of *GmZOU*s in Vegetatives

The expression profiles of the *GmZOU*s in both vegetative and reproductive soybean organs, including roots, stems, leaves, inflorescences, unopened flowers, opened flowers and developing seeds with embryos in different stages were examined by qPCR analysis. The results showed that both *GmZOU*s were highly expressed in developing seeds and preferentially at the heart stage. *GmZOU-2* was also detected in the inflorescences, unopened flowers, and opened flowers (Figure 3A).

In order to study the spatial-specific expression pattern of *GmZOU*s in developing seeds, gene expression analysis had been carried out on dissected seed compartments. The embryo and

endosperm were separately isolated from developing seeds at late heart stage or cotyledon stage. In order to prevent contamination, the expression of embryo specific genes (*Glyma03g03500* and *Glyma14g35560*) and endosperm specific genes (*Glyma07g02220* and *Glyma08g21890*; Danzer et al., 2015) was also checked in these dissected seed compartments (Supplementary Figure S2). Upon verification, *GmZOU*s expression was detected by qPCR analysis. The results showed that both of *GmZOU*s were highly expressed in endosperm whereas no expression was detected in embryo (Figure 3B). The similar expression pattern with *AtZOU* implied functional conservation of *GmZOU*s in seed development.

The Expression of *GmZOU*s in the *Arabidopsis zou* Mutant Facilitates the Recovery of the Mutant Seed Phenotype

To confirm that *GmZOU*s are involved in endosperm breakdown, expression vectors with *GmZOU-1* and *GmZOU-2* under the control of the *AtZOU* promoter (*AtZOUpro-GmZOU*s) were separately transformed into *Atzou-4* mutants. After resistance

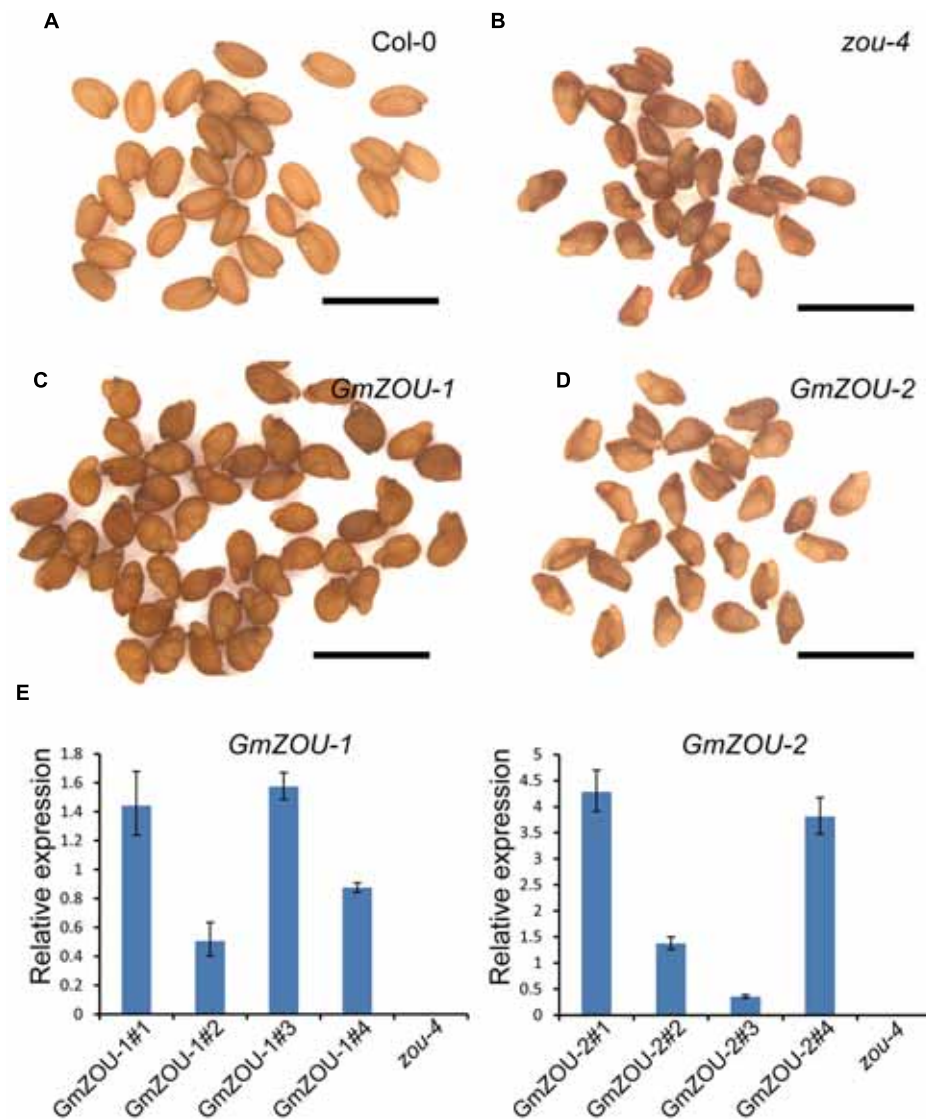


FIGURE 4 | *GmZOUS* expressed in *Arabidopsis zou-4* mutant. (A–D) Seed phenotype of *GmZOU-1* and *GmZOU-2* transgenic lines. The seeds from Col-0 are full, with the embryo showing a perfectly bent shape (A) and from *zou-4* are abnormal with a shriveled seed phenotype (B). In *GmZOU-1* transgenic lines, the seeds showed a wild-type-like or malformed seed phenotype (C), whereas the seeds in the *GmZOU-2* transgenic lines showed the shriveled seed phenotype, similar to that in the *Atzou-4* mutant (D). Scale bars: 1 mm. (E) qPCR analysis of the *GmZOUS* expression in transgenic lines. Four independent lines from each were selected for examining the exogenous *GmZOUS* expression, which was quantified relative to *AtZOU* expression in Col-0 wild type seeds at the heart stage. The un-transformed *Atzou-4* mutants were selected as a negative control.

screening, presence of exogenous *AtZOU* promoter-driven *GmZOU-1* or *GmZOU-2* constructs was checked in the transgenic lines containing the T-DNA insertions. The *Atzou-4* mutant background of the *GmZOU-1* and *GmZOU-2* transgenic lines was then confirmed by PCR analysis based on the T-DNA insertion site. Furthermore, the expression of exogenous *GmZOUS* was also detected by qPCR (Figure 4E). Background identification and exogenous *GmZOUS* expression analysis identified a total of four *GmZOU-1* transgenic lines and four *GmZOU-2* transgenic lines, which were then used in the subsequent analysis. Among these transgenic lines, the mature seeds in the *GmZOU-1* transgenic lines rescued the *Atzou-4* shriveled seed phenotype

into wild-type-like or malformed seeds (Figures 4A,C), whereas the mature seeds in the *GmZOU-2* transgenic lines showed no distinct phenotypic differences from that of *Atzou-4* mutant seeds (Figures 4B,D).

To study the embryo and endosperm growth in these transgenic lines, the developing seeds at about 10–12 days after pollination (DAP) were observed. The Col-0 wild-type embryos were in the mature stage showing a completely bent embryo, and the endosperm had already undergone complete breakdown with only one cell layer surrounding the embryo (Figures 5A,F); the embryo growth in *Atzou-4* mutant was restricted at the heart/torpedo shaped embryo, and the endosperm failed to

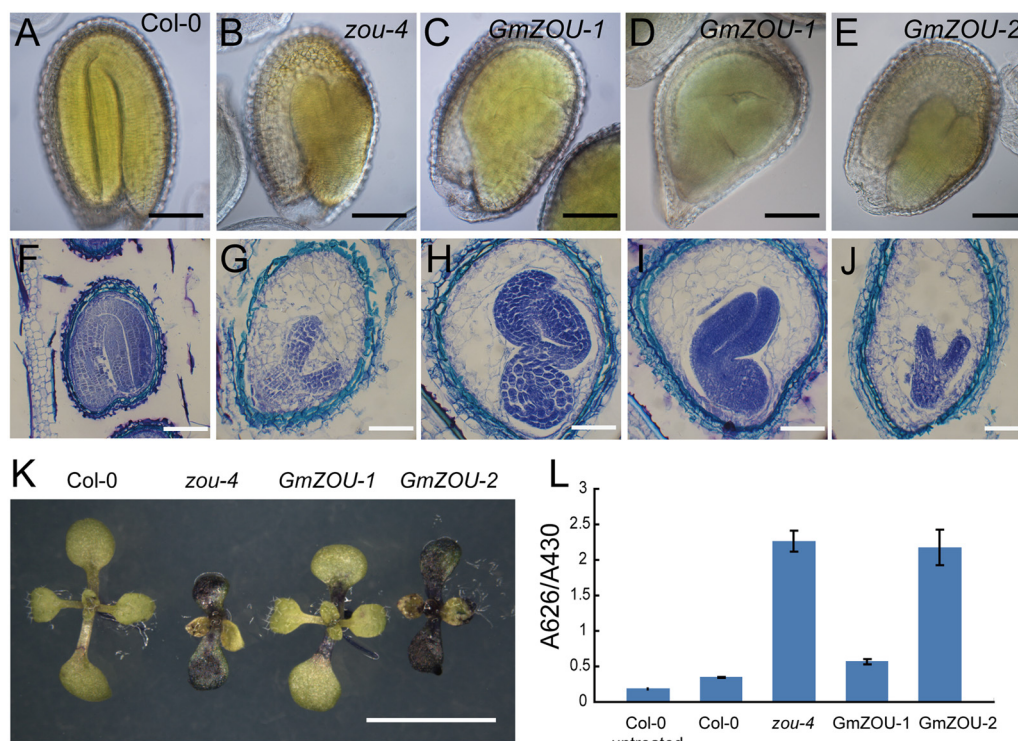


FIGURE 5 | Evaluation of seed development and detection of the embryo cuticle. (A–E) The seeds at 12 DAP were cleared and observed by DIC microscopy. The embryo in Col-0 showed complete bending, occupying the entire seed cavity (**A**) and arrested in the heart/torpedo stage in the *Atzou-4* mutant (**B**). In *GmZOU-1* transgenic seeds, the semi-bent and larger embryo occupied the seed (**C,D**), and the embryo in the *GmZOU-2* transgenic seeds was the same as that in the *Atzou-4* mutant (**E**). Scale bars: 100 μ m. (**F–J**) Embryonic and endosperm development was detected by paraffin sectioning at about 10 DAP. The wild-type seeds were in the mature stage, showing a completely bent embryo and only one cell layer of endosperm surrounding the embryo (**F**); the *Atzou-4* mutant embryos were arrested in the heart stage and their cellularized endosperm remained intact (**G**). In *GmZOU-1* transgenic seeds, the endosperm showed breakdown and the embryo expanded to later stages. Furthermore, the expanded embryo still exhibited non-uniform development based on the amount of residual endosperm (**H,I**). In the *GmZOU-2* transgenic seeds, the embryo and endosperm remained arrested, similar to that in the *Atzou-4* mutant (**J**). Scale bars: 100 μ m in (**F**) and 50 μ m in (**G–J**). (**K,L**) *GmZOU-1* partially recovered the cuticle deficiency in *Atzou-4*. The seedlings of Col-0, *Atzou-4*, and the *GmZOU*s transgenic lines at about 7 days after germination were stained with TB. The *GmZOU-1* transgenic seedlings were weakly stained, whereas the *GmZOU-2* transgenic seedlings showed strong staining similar to that in the *Atzou-4* mutants (**K**). TB uptake in these lines was quantified spectrophotometrically, showing similar results to that of microscopy analysis. Error bars represent s.d. among three biological replicates, each containing 20 seedlings (**L**). Scale bars: 5 mm.

undergo a breakdown, with cellularized endosperm persisting (Figures 5B,G). The seeds of the *GmZOU-2* transgenic lines showed similar features as that of the *Atzou-4* mutant, with an embryo arrested in the heart/torpedo stage and showing a sustained cellularized endosperm (Figures 5E,J). On the contrary, the embryo of the *GmZOU-1* transgenic lines showed elongated and bending, with the semi-bent embryo almost occupying the entire embryonic sac (Figures 5C,D); and the endosperm showed breakdown in *GmZOU-1* transgenic lines resulting in a reduction in the amount of persistent endosperm compared with *Atzou-4* mutant (Figures 5H,I). The findings of seed development observation confirmed that the *GmZOU-1* partially recovered embryo expansion and endosperm breakdown in *Atzou-4* mutants, whereas *GmZOU-2* did not influence *Atzou-4* mutant seed development.

Furthermore, we also studied the embryonic cuticle in the transgenic lines as the cuticle integrity in *zou* was not intact with breaks in epidermal. TB, a hydrophilic dye, was used to detect the hydrophobic cuticle and epidermal defection

(Tanaka et al., 2004). The seedlings of Col-0, *Atzou-4*, and the *GmZOU*s transgenic lines at about 7 days after germination were stained with TB and visualized under a dissecting microscope. Cotyledons of the *Atzou-4* mutants showed strong staining, while the cotyledons of Col-0 wild-type had no visible permeability to TB. The transgenic *GmZOU-2* cotyledons also showed strong staining similar to that in the *Atzou-4* mutants, whereas the *GmZOU-1* transgenic lines were weakly stained (Figure 5K). To further investigate this phenotype, TB uptake by seedlings was quantified by using a spectrophotometer. The results were similar to that of microscopy analysis (Figure 5L). These findings indicate that the *GmZOU-1* transgenic plants also partially rescued cuticle formation, whereas *GmZOU-2* did not.

***GmZOU-1* Recovered the Expression of Presumed ZOU Target Genes**

The expression of presumed ZOU target genes, including *ALE1*, *At3g06890*, *At5g03820*, and *At4g33600* (Kondou et al., 2008;

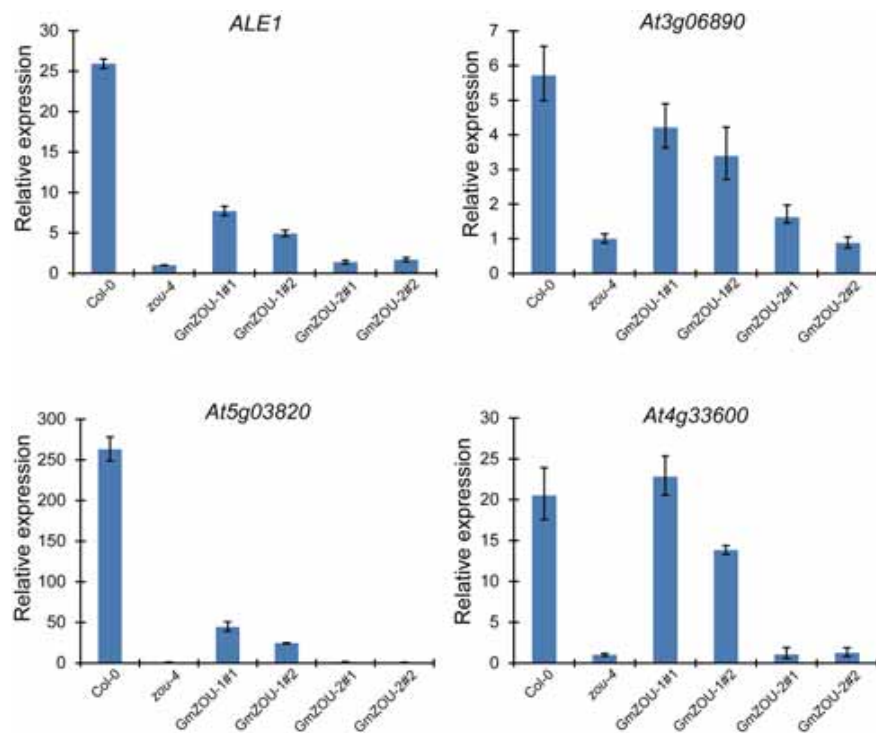


FIGURE 6 | The expression of AtZOU regulated genes in Col-0, *Atzou-4*, *GmZOU-1*, and *GmZOU-2* transgenic lines as detected by qPCR. The expression of *ALE1*, *At3g06890*, *At5g03820*, and *At4g33600*, which are the AtZOU presumed target genes, decreased in the *Atzou-4* mutant and showed partial recovery in the *GmZOU-1* transgenic lines, but not in the *GmZOU-2* transgenic lines.

Yang et al., 2008; Xing et al., 2013) were detected in the *GmZOU*s transgenic lines (Figure 6). The ZOU target genes *At3g06890* and *At4g33600* showed almost full recovery in the *GmZOU-1* transgenic lines. The mRNA levels of the putative direct target *ALE1* (Yang et al., 2008; Xing et al., 2013) increased sevenfold and threefold in the *GmZOU-1* transgenic lines compared with levels observed in untransformed *Atzou-4* mutants but reached only 30 and 13%, respectively, of the level observed in Col-0. Similarly, the expression of target gene *At5g03820* in the *GmZOU-1* transgenic lines reached only 17 and 9.3%, respectively, relative to that in Col-0. On the other hand, the expression of these target genes did not differ in the *GmZOU-2* transgenic lines compared to that in the *Atzou-4* mutant.

These results showed that the downregulated genes in the *Atzou-4* mutant underwent partial recovery in the *GmZOU-1* transgenic lines, whereas these did not change in the *GmZOU-2* transgenic lines. In the *GmZOU-1* transgenic lines, the partially recovered expression of these target genes explained the observed partial recovery of the *Atzou-4* mutant phenotype. The less efficient recovery of the ZOU target genes such as *ALE1* and *At5g03820* implied that part of the ZOU target genes requires higher expression levels or complete ZOU function. Nonetheless, our data confirmed that *GmZOU-1* is a functional ortholog of AtZOU because both soybean and *Arabidopsis* ZOU induced the upregulation of ZOU presumed target genes, which have been associated with endosperm breakdown and embryo cuticle development.

DISCUSSION

GmZOU-1 Is the ZOU Ortholog in Soybean

Searching the soybean genomic sequence database identified two AtZOU homologs in soybean, which were then further confirmed by sequence alignment and chromosome synteny analysis. The results also confirmed that the AtZOU homologs resulted from tandem duplication. To confirm functional conservation, the *GmZOU*s were expressed in *Arabidopsis* *Atzou-4* mutants and only *GmZOU-1* could complement the *Atzou-4* mutant phenotype with endosperm breakdown, embryo expansion, and cuticle formation. Additionally, the homologs showed expression patterns that were similar to that of AtZOU, which indicated that *GmZOU-1* is the ZOU ortholog in soybean. *GmZOU-1* only partially complemented the *Atzou-4* mutant, which was indicative of species specificity or functional diversity during evolution.

GmZOU-2 Is Derived from Gene Duplication and Lost Function during the Course of Evolution

GmZOU-1 (*Glyma.02G103200*) and *GmZOU-2* (*Glyma.02G103100*) resulted from tandem duplication with the same transcription orientation. Sequence alignment indicated that *GmZOU-1* consisted of a classical bHLH DNA binding domain

at the N terminus and a conserved functional domain at the C terminus compared to AtZOU. By contrast GmZOU-2 only had the conserved function domain, whereas the bHLH domain had lost the conserved amino acids which may influence its DNA binding capability. Transgenic analysis indicated that *GmZOU-2* did not influence the *Atzou-4* mutant in terms of embryonic arrest at the heart stage, sustaining the endosperm, and cuticle formation deficiency. Additionally, considering its less specific expression pattern, we concluded that *GmZOU-2* is the product of gene duplication and may have lost its original function during the course of evolution.

Cuticle Deficiency and Arrested Endosperm Breakdown Result in Embryonic Malformation in *GmZOU-1* Transgenic Lines

The cuticularization of juxtaposed surfaces has been shown to be extremely important in defining organ boundaries. Mutants with compromised cuticles often show extensive organ fusions (Kurdyukov et al., 2006; Xing et al., 2013). In the *GmZOU-1* transgenic lines, *ALE1* expression was partially recovered, and TB staining showed that the cuticle layer of the embryonic cotyledons did not show complete recovery. The embryo with the defective cuticle readily fused with the endosperm, similar to that observed in other embryo cuticle-deficient mutants such as *ale1*, *gso1/gso2*, and *zou* (Tanaka et al., 2001; Tsuwamoto et al., 2008; Yang et al., 2008). We inferred that the observed organ fusion suppressed embryo expansion. The combination of organ fusion in the embryo and the arrest of endosperm degradation restricted embryonic development, which resulted in the reverse bending of the embryo and a severe malformation phenotype (Figure 5). These mechanisms may explain the development of malformed embryos in the *GmZOU-1* transgenic seeds.

REFERENCES

- Abbadi, A., and Leckband, G. (2011). Rapeseed breeding for oil content, quality, and sustainability. *Eur. J. Lipid Sci. Technol.* 113, 1198–1206. doi: 10.1002/ejlt.201100063
- Baud, S., Dubreucq, B., Miquel, M., Rochat, C., and Lepiniec, L. (2008). Storage reserve accumulation in *Arabidopsis*: metabolic and developmental control of seed filling. *Arabidopsis Book* 6:e0113. doi: 10.1199/tab.0113
- Baud, S., Wuielleme, S., Lemoine, R., Kronenberger, J., Caboche, M., Lepiniec, L., et al. (2005). The AtSUC5 sucrose transporter specifically expressed in the endosperm is involved in early seed development in *Arabidopsis*. *Plant J.* 43, 824–836. doi: 10.1111/j.1365-313X.2005.02496.x
- Berger, F. (2003). Endosperm: the crossroad of seed development. *Curr. Opin. Plant Biol.* 6, 42–50. doi: 10.1016/S1369526602000043
- Berger, F., Grini, P. E., and Schnittger, A. (2006). Endosperm: an integrator of seed growth and development. *Curr. Opin. Plant Biol.* 9, 664–670. doi: 10.1016/j.pbi.2006.09.015
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313X.1998.00343.x
- Danzer, J., Mellott, E., Bui, A. Q., Le, B. H., Martin, P., Hashimoto, M., et al. (2015). Down-regulating the expression of 53 soybean transcription factor genes uncovers a role for SPEECHLESS in initiating stomatal cell lineages during embryo development. *Plant Physiol.* 168, 1025–1035. doi: 10.1104/pp.15.00432

AUTHOR CONTRIBUTIONS

SY and XF conceived the project and designed this work. YZ cloned the *GmZOU*s and performed phylogenetic and gene expression analyses. YZ and XL performed transgenic, cell biological, and other functional analyses. SY and XF wrote the manuscript.

FUNDING

This work was supported by the National Nature Science Foundation of China (grant nos. 31470286 and 31571692), the One Hundred Person Project of the Chinese Academy of Sciences, and the Strategic Priority Research Program of the Chinese Academy of Sciences (grant no. XDA0801010502) and was also supported by the National Key Research and Development Project (grant no. 2016YFD0101900) from the Ministry of Science and Technology of China.

ACKNOWLEDGMENT

We thank Accdon for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00139/full#supplementary-material>

- D'Erfurth, I., Le Signor, C., Aubert, G., Sanchez, M., Vernoud, V., Darchy, B., et al. (2012). A role for an endosperm-localized subtilase in the control of seed size in legumes. *New Phytol.* 196, 738–751. doi: 10.1111/j.1469-8137.2012.04296.x
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678
- Grimault, A., Gendrot, G., Chamot, S., Widiez, T., Rabillé, H., Gérentes, M. F., et al. (2015). ZmZHOUP1, an endosperm-specific bHLH transcription factor involved in maize seed development. *Plant J.* 84, 574–586. doi: 10.1111/tj.13024
- Hill, L. M., Morley-Smith, E. R., and Rawsthorne, S. (2003). Metabolism of sugars in the endosperm of developing seeds of oilseed rape. *Plant Physiol.* 131, 228–236. doi: 10.1104/pp.010868
- Hou, A., Chen, P., Alloatti, J., Li, D., Mozzoni, L., Zhang, B., et al. (2009). Genetic variability of seed sugar content in worldwide soybean germplasm collections. *Crop Sci.* 49, 903–912. doi: 10.2135/cropsci2008.05.0256
- Hu, R., Fan, C., Li, H., Zhang, Q., and Fu, Y. F. (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol. Biol.* 10:93. doi: 10.1186/1471-2199-10-93
- Ingram, G. C. (2010). Family life at close quarters: communication and constraint in angiosperm seed development. *Protoplasma* 247, 195–214. doi: 10.1007/s00709-010-0184-y
- Jones, S. I., and Vodkin, L. O. (2013). Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS ONE* 8:e59270. doi: 10.1371/journal.pone.0059270
- Kondou, Y., Nakazawa, M., Kawashima, M., Ichikawa, T., Yoshizumi, T., Suzuki, K., et al. (2008). RETARDED GROWTH OF EMBRYO1, a

- new basic helix-loop-helix protein expresses in endosperm to control embryo growth. *Plant Physiol.* 147, 1924–1935. doi: 10.1104/pp.108.118364
- Kurdyukov, S., Faust, A., Nawrath, C., Bär, S., Voisin, D., Efremova, N., et al. (2006). The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*. *Plant Cell* 18, 321–339. doi: 10.1105/tpc.105.036079
- Luo, M., Dennis, E. S., Berger, F., Peacock, W. J., and Chaudhury, A. (2005). MINISEED3 (MINI3), a WRKY family gene, and HAIKU2 (IKU2), a leucine-rich repeat (LRR) KINASE gene, are regulators of seed size in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17531–17536. doi: 10.1073/pnas.0508418102
- Melkus, G., Rolletschek, H., Radchuk, R., Fuchs, J., Rutten, T., Wobus, U., et al. (2009). The metabolic role of the legume endosperm: a noninvasive imaging study. *Plant Physiol.* 151, 1139–1154. doi: 10.1104/pp.109.143974
- Olsen, O. A. (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* 16, S214–S227. doi: 10.1105/tpc.017111
- Otegui, M. S., Capp, R., and Staehelin, L. A. (2002). Developing seeds of *Arabidopsis* store different minerals in two types of vacuoles and in the endoplasmic reticulum. *Plant Cell* 14, 1311–1327. doi: 10.1105/tpc.010486
- Rosso, M. G., Li, Y., Strizhov, N., Reiss, B., Dekker, K., and Weisshaar, B. (2003). An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. *Plant Mol. Biol.* 53, 247–259. doi: 10.1023/B:PLAN.0000009297.37235.4a
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Schwender, J., and Ohlrogge, J. B. (2002). Probing in vivo metabolism by stable isotope labeling of storage lipids and proteins in developing *Brassica napus* embryos. *Plant Physiol.* 130, 347–361. doi: 10.1104/pp.004275
- Sharma, V. K., Ramirez, J., and Fletcher, J. C. (2003). The *Arabidopsis* CLV3-like (CLE) genes are expressed in diverse tissues and encode secreted proteins. *Plant Mol. Biol.* 51, 415–425. doi: 10.1023/A:1022038932376
- Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* 24, 1596–1599. doi: 10.1093/molbev/msm092
- Tanaka, H., Onouchi, H., Kondo, M., Hara-Nishimura, I., Nishimura, M., Machida, C., and Machida, Y. (2001). A subtilisin-like serine protease is required for epidermal surface formation in *Arabidopsis* embryos and juvenile plants. *Development* 128, 4681–4689.
- Tanaka, T., Tanaka, H., Machida, C., Watanabel, M., and Machida, Y. (2004). A new method for rapid visualization of defects in leaf cuticle reveals five intrinsic patterns of surface defects in *Arabidopsis*. *Plant J.* 37, 139–146. doi: 10.1046/j.1365-313X.2003.01946.x
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. doi: 10.1093/nar/22.22.4673
- Tsuwamoto, R., Fukuoka, H., and Takahata, Y. (2008). GASSHO1 and GASSHO2 encoding a putative leucine-rich repeat transmembrane-type receptor kinase are essential for the normal development of the epidermal surface in *Arabidopsis* embryos. *Plant J.* 54, 30–42. doi: 10.1111/j.1365-313X.2007.03395.x
- Xing, Q., Creff, A., Waters, A., Tanaka, H., Goodrich, J., and Ingram, G. C. (2013). ZHOUP1 controls embryonic cuticle formation via a signalling pathway involving the subtilisin protease ABNORMAL LEAFSHAPE1 and the receptor kinases GASSHO1 and GASSHO2. *Development* 140, 770–779. doi: 10.1242/dev.088898
- Yang, S., Johnston, N., Talideh, E., Mitchell, S., Jeffree, C., Goodrich, J., et al. (2008). The endosperm-specific ZHOUP1 gene of *Arabidopsis thaliana* regulates endosperm breakdown and embryonic epidermal development. *Development* 135, 3501–3509. doi: 10.1242/dev.026708

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Zhang, Li, Yang and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Conserved Potential Development Framework Applies to Shoots of Legume Species with Contrasting Morphogenetic Strategies

Lucas Faverjon, Abraham J. Escobar-Gutiérrez, Isabelle Litrico and Gaëtan Louarn *

INRA, UR4, URP3F, BP6, Lusignan, France

OPEN ACCESS

Edited by:

Maria Carlota Vaz Patto,
Instituto de Tecnologia Química e
Biológica, Universidade Nova de
Lisboa, Portugal

Reviewed by:

Hao Peng,
Washington State University, USA
Gerda Cnops,
Institute for Agricultural and Fisheries
Research (ILVO), Belgium

*Correspondence:

Gaëtan Louarn
gaetan.louarn@inra.fr

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 30 November 2016

Accepted: 09 March 2017

Published: 27 March 2017

Citation:

Faverjon L, Escobar-Gutiérrez AJ,
Litrico I and Louarn G (2017) A
Conserved Potential Development
Framework Applies to Shoots of
Legume Species with Contrasting
Morphogenetic Strategies.
Front. Plant Sci. 8:405.
doi: 10.3389/fpls.2017.00405

A great variety of legume species are used for forage production and grown in multi-species grasslands. Despite their close phylogenetic relationship, they display a broad range of morphologies that markedly affect their competitive abilities and persistence in mixtures. Little is yet known about the component traits that control the deployment of plant architecture in most of these species. During the present study, we compared the patterns of shoot organogenesis and shoot organ growth in contrasting forage species belonging to the four morphogenetic groups previously identified in herbaceous legumes (i.e., stolon-formers, rhizome-formers, crown-formers tolerant to defoliation and crown-formers intolerant to defoliation). To achieve this, three greenhouse experiments were carried out using plant species from each group (namely alfalfa, birdsfoot trefoil, sainfoin, kura clover, red clover, and white clover) which were grown at low density under non-limiting water and soil nutrient availability. The potential morphogenesis of shoots characterized under these conditions showed that all the species shared a number of common morphogenetic features. All complied with a generalized classification of shoot axes into three types (main axis, primary and secondary axes). A common quantitative framework for vegetative growth and development involved: (i) the regular development of all shoot axes in thermal time and a deterministic branching pattern in the absence of stress; (ii) a temporal coordination of organ growth at the phytomer level that was highly conserved irrespective of phytomer position, and (iii) an identical allometry determining the surface area of all the leaves. The species differed in their architecture as a consequence of the values taken by component traits of morphogenesis. Assessing the relationships between the traits studied showed that these species were distinct from each other along two main PCA axes which explained 68% of total variance: the first axis captured a trade-off between maximum leaf size and the ability to produce numerous phytomers, while the second distinguished morphogenetic strategies reliant on either petiole or internode expansion to achieve space colonization. The consequences of this quantitative framework are discussed, along with its possible applications regarding plant phenotyping and modeling.

Keywords: forage legumes, morphogenesis, branching, architecture, leaf area, growth habit, competitive ability

INTRODUCTION

Numerous forage legumes contribute to temperate grasslands and help to supply high-quality protein-rich feed for ruminants, while reducing the need for nitrogen fertilizers (Suter et al., 2015; Vertès et al., 2015), preserving water quality (Owens et al., 1994; Russelle et al., 2001) and mitigating greenhouse gas emissions (Jensen et al., 2012). Most of these legume species are grown in a mixture with perennial grasses in order to take advantage of the ecological and nutritional complementarities of the two functional groups (Nyfeler et al., 2011; Gaba et al., 2015). However, a long-acknowledged problem of multi-species grasslands is the lack of persistence of the legume component over time and a less predictable forage quality when compared with pure grasses or annual forages (Beuselinck et al., 1994; Schwinning and Parsons, 1996). Competition for resources and crop management have been shown to be of considerable importance to regulating the proportion of legumes in grassland communities (Sheaffer, 1989; Beuselinck et al., 1994), but little is known about the mechanisms by which a legume species prevails within a particular community or environment. To date, the search for combinations of traits predictive of legume performance in a mixture has been limited to a few widely-grown mixtures (e.g., white clover-perennial ryegrass or alfalfa-tall fescue: Davies, 2001; Annicchiarico et al., 2015; Maamouri et al., 2015), and only a few studies have sought to explore the role of the diversity of plant forms using multi-trait approaches (e.g., Fort et al., 2015 on root traits; Kraft et al., 2015).

The legume family presents spectacular morphological and life-history trait diversity (Lewis et al., 2005; LPWG, 2013). Considering the temperate herbaceous genus only, a remarkable range of plant structures and pedoclimatic adaptations has been reported (Forde et al., 1989; Scott et al., 1989). The morphogenesis of shoots determines plant architecture and light capture (Valladares and Niinemets, 2007), which are critical to inferring the outcome of competition for light (Caldwell, 1987). It also determines the position of shoot meristems and contributes to how different species tolerate grazing and mowing (Briske, 1996; Smith et al., 2000). Differences in leaf surface area, plant height and other architectural features that affect the spatial distribution of leaves (e.g., phenology, branching patterns, dimensions of spacing organs, etc.) have been shown to be drivers of competitive success in grass-legume mixtures (Louarn et al., 2012; Barillot et al., 2014). However, little is known about the elementary traits that promote leaf area expansion and height increments in most herbaceous species, especially regarding the temporal aspects of morphogenesis. Unlike grasses, for which a regular developmental scheme was identified and mobilized a long time ago to compare species and genotypes (Simon and Lemaire, 1987; Lafarge and Durand, 2013), no obvious pattern has emerged from comparisons of shoots from crown-, stolon- and rhizome-forming legumes used for forage production (Forde et al., 1989; Thomas, 2003; **Figure 1**).

As parts of a modular organism, the common feature of all plant shoots is that they are built from elementary subunits called phytomers (White, 1979). Shoot morphogenesis thus arises from the initiation of new phytomers by shoot meristems, from

the expansive growth of the individual organs produced and from the differentiation of support tissues. Like many other dicotyledonous plants (Seleznyova et al., 2002; Lebon et al., 2006), forage legumes are characterized by shoots with complex and highly branched structures (e.g., white clover: Thomas, 1987; Gautier et al., 2000; Thomas et al., 2014; red clover: Taylor and Quesenberry, 1996). Considerable variability of shoot development is usually observed in dense stands (Gosse et al., 1988; Van Minnebruggen et al., 2015). However, architectural and developmental analyses have proved to be powerful tools to classify shoot types and quantify branching and potential shoot development in different species and varieties, based on stochastic (Costes and Guédon, 2002; Louarn et al., 2007) and deterministic (Turc and Lecoecur, 1997) approaches. Although, using different methods, regular organogenetic patterns have been found in alfalfa (Baldissera et al., 2014), red clover (Van Minnebruggen et al., 2014), and *Medicago truncatula* (Moreau et al., 2006), suggesting that a generalized description could be envisioned for these species.

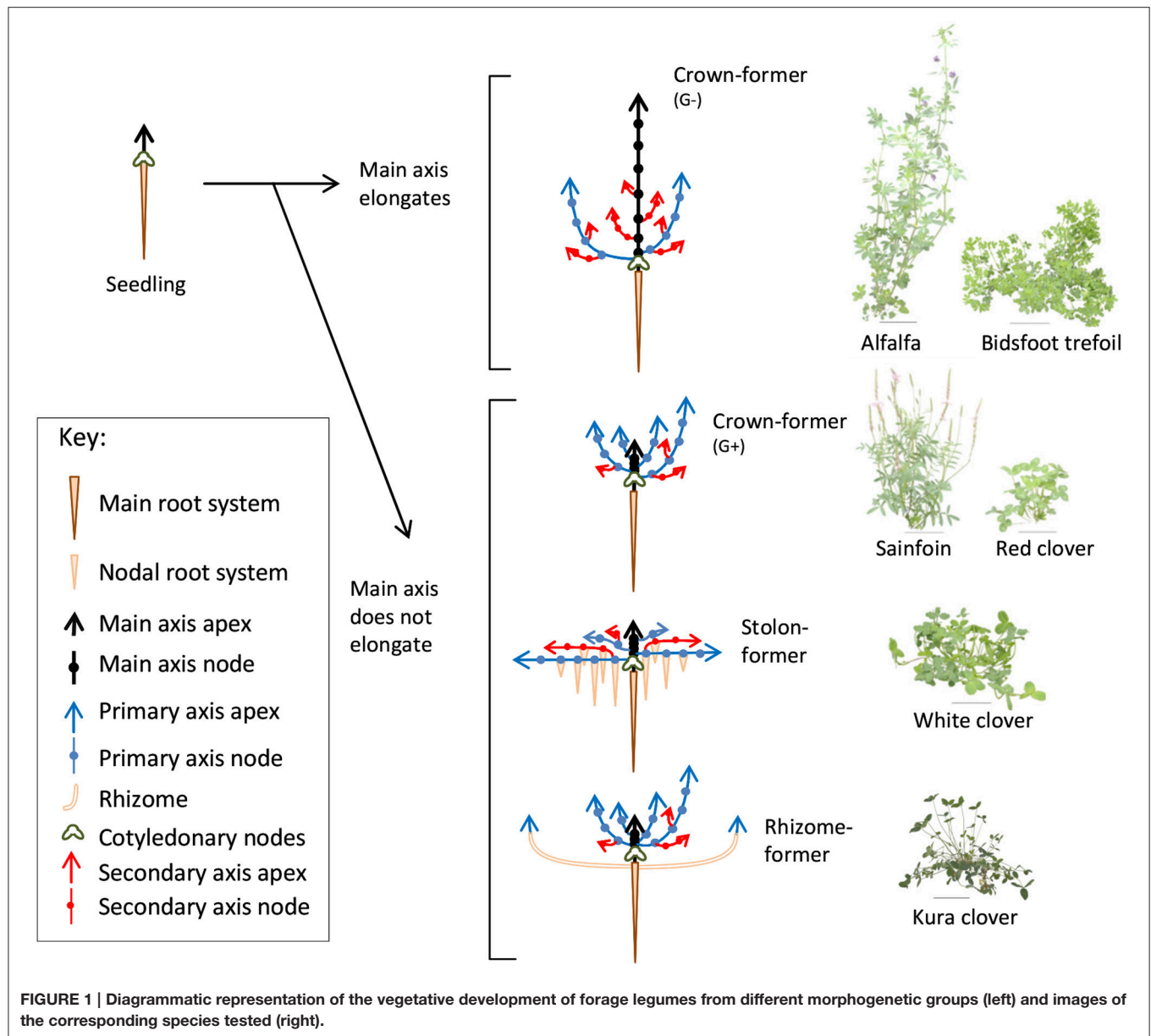
Similarly, the complexity of plant growth, distributed between individual organs that all vary in size and shape, has been shown to be driven by a number of common determinants. For instance, the temporal sequences of organ growth generally appear to be coordinated at a supra-organ level. In grasses, changes to the phases of leaf growth are triggered by emergence of the previous leaves (Skinner and Nelson, 1994; Fournier et al., 2005; Louarn et al., 2010). In many other species, stable thermal time calendars of development have been reported at the plant and phytomer levels in the absence of stress (Granier et al., 2002; Lecoecur, 2010; Demotes-Mainard et al., 2013). Furthermore, the ultimate sizes attained by organs are highly heritable traits (Annicchiarico et al., 1999), which present relatively conserved ontogenic patterns along the stem under controlled conditions (Ross, 1981). Overall, organogenetic and expansive growth characteristics define and—by their ongoing interactions constrain—the shoot morphogenesis of any particular plant genotype. Trade-offs very often occur (e.g., in grasses between leaf growth and tillering; Nelson, 2000; Barre et al., 2015), which therefore makes it worthwhile analysing these two aspects of morphogenesis together.

In order to facilitate the future characterization of shoot morphogenesis in different legume species and cultivars, this paper was designed to: (i) analyse the elementary traits controlling plant leaf area and height in a selection of six herbaceous species, contrasting in terms of their growth habits and architectures, (ii) challenge the existence of a common framework for vegetative development under non-limiting growing conditions, and (iii) assess the relationships between the traits studied and the concomitant occurrence of trait values which might indicate ontogenic trade-offs in the morphogenesis of shoots.

MATERIALS AND METHODS

Plant Materials and Growing Conditions

Three experiments were carried out in a greenhouse at INRA Lusignan, France (46°26' N, 0°07' E). The two first took place



from February to April in 2014 and 2015, whereas the third experiment was carried out from November 2016 to February 2017. Plants of alfalfa (*Medicago sativa* L. cv. Timbale; hereinafter referred to as A), white clover (*Trifolium repens* L. cv. Giga; WC), red clover (*Trifolium pratense* L. cv. Formica; RC), and sainfoin (*Onobrychis viciifolia* Scop. cv. Canto; SF) were studied during the two first experiments. In addition, plants of birdsfoot trefoil (*Lotus corniculatus* L. cv. Leo; BT) and kura clover (*Trifolium ambiguum* cv. Sevanskij; KC) were grown in the second and third years. Overall, the six species selected covered a wide range of shoot growth habits and the four morphogenetic groups previously reported among perennial herbaceous legumes from temperate areas (Forde et al., 1989; Thomas, 2003; **Figure 1**) and adapted to contrasting ecological niches (Scott et al., 1989).

In each experiment, seeds from each species were germinated for 48 h in Petri dishes at 25°C in the dark. In addition, clones propagated from rhizome cuttings were selected on 2-years-old plants and used for TK in the third experiment. The seedlings were planted 0.3 m × 1 m apart in custom-built 10 L boxes containing sterile potting mix, sand and brown soil (1:1:1, v/v/v). For each species, the plants were grown in isolation (3.3 plant.m⁻²) until the end of the vegetative stage, according to a randomized block design with four (Exp. 1 and 2) to five (Exp. 3) replicates by species. Irrigation and fertilization were provided throughout the experiments with a drip system 5 cm distant from the plants that delivered 300 mL.d⁻¹ of complete nutrient solution (Gastal and Saugier, 1986). The nitrogen concentration of the solution (8 mM) was non-limiting for growth and

prevented the nodulation of roots in all the legume species. The greenhouse was heated and a photoperiod of 16 h was maintained by means of 400-Watt HQI lamps (Supplementary Table S1).

Plant Measurements

To analyse the temporal development of photosynthetic surfaces and their spatial distributions in each species, the organogenesis and expansion of shoot organs were measured for all the plants for a period of about 70 days. The terminology used to refer to the different shoot axes was based on the architectural descriptions of alfalfa and white clover (Mouliat et al., 2000; Baldissera et al., 2014) and is summarized in Supplementary Figure S1A.

The numbers of axes and numbers of phytomers on the different axes were counted weekly. A decimal scale was used to account for phytomers with unfolded leaves (Maitre et al., 1985). Furthermore, several groups of three consecutive phytomers were selected for daily growth measurements on both primary and secondary axes. Ranks 5, 6, 7 and 11, 12, 13 were followed on the main axis. Ranks 3, 4, 5 were characterized on primary axes and branches. The length of each organ (i.e., leaflet, petiole, internode in all species; stipule in birdsfoot trefoil; Supplementary Figure S1B) was measured every day with a ruler until no further growth was noted over 4 consecutive days.

At the end of each experiment, the final length and width of each mature organ along the primary axes was measured. In experiments 1 and 2, a sub-sample of phytomers was used in each species to determine the leaf area of individual organs. Leaves and stipules of various sizes and positions were scanned (Konica Minolta C352/C300, Konica Minolta Sensing, Osaka, Japan) and their area was measured by image analysis (ImageJ software, <http://rsbweb.nih.gov/ij/>). The height of plants, the total plant leaf area and the total dry weight of plants were also measured.

Meteorological Measurements and Thermal Time Calculations

Relative humidity was measured using a capacitive hygrometer (HMP35A Vaisala, Oy, Helsinki, Finland) and air temperature with copper-constantan thermocouples placed in a ventilated radiation shield at the center of the greenhouse. Photosynthetic photon flux density (PPFD) was also measured by means of PPFD sensors placed above each experimental bloc within the greenhouse. All data were stored in a datalogger (CR10X, Campbell Scientific Ltd.), with measurements taken every 30 s and an average calculated over 15 min. The data are summarized for each experiment in Supplementary Table S1.

Thermal time (TT) was calculated as the integral of a non-linear beta function of temperature (T) as proposed by Zaka et al. (2017):

$$f(T) = \left(\frac{T - T_{\min}}{T_{\text{ref}} - T_{\min}} \right)^q \cdot \left(\frac{T_{\max} - T}{T_{\max} - T_{\text{ref}}} \right) \quad (1)$$

$$TT = \int_{t_0}^t \max[0, (T_{\text{ref}} - T_{\text{base}}) \cdot f(T)] \cdot dt \quad (2)$$

The equation has three parameters: the minimum (T_{\min}) and maximum (T_{\max}) temperatures at which development occurs

and q , a shape parameter. In addition, T_{ref} accounts for a fixed reference temperature (20°C) and T_{base} (5°C) for a common base temperature used to scale time units from equivalent days at the reference temperature into degree days ($^{\circ}\text{Cd}$). The parameters used for the different species were derived from the literature and are presented in Supplementary Table S2.

Data Analysis

All calculations and statistical tests were performed using R software (version 3.1.2; R Development Core Team, 2014). Rates of leaf appearance were calculated for each axis with three or more unfolded leaves by linear regression between thermal time and the number of visible phytomers. Phyllochrons were calculated as the reciprocal of the leaf appearance rate.

The temporal growth of plant organs was analyzed using a three-parameter logistic function (Equation 3) fitted to the time series of organ growth measurements:

$$L(t)/L_{\max} = \frac{1}{1 + e^{-s \cdot (t - t50)}} \quad (3)$$

where the s and $t50$ parameters represent the steepness and time delay at mid-organ expansion and L_{\max} the final organ length. By convention, all the organs within a phytomer were analyzed with respect to the leaf appearance (i.e., $t50$ of leaflets = 0) and time was expressed in phyllochrons to aggregate growth series from axes with different developmental rates. The duration of organ expansion between 5 and 95% of its final dimension (d_{95}) was derived for each organ from the s parameter as follow:

$$d_{95} = -2 \cdot \ln(0.05/0.95)/s \quad (4)$$

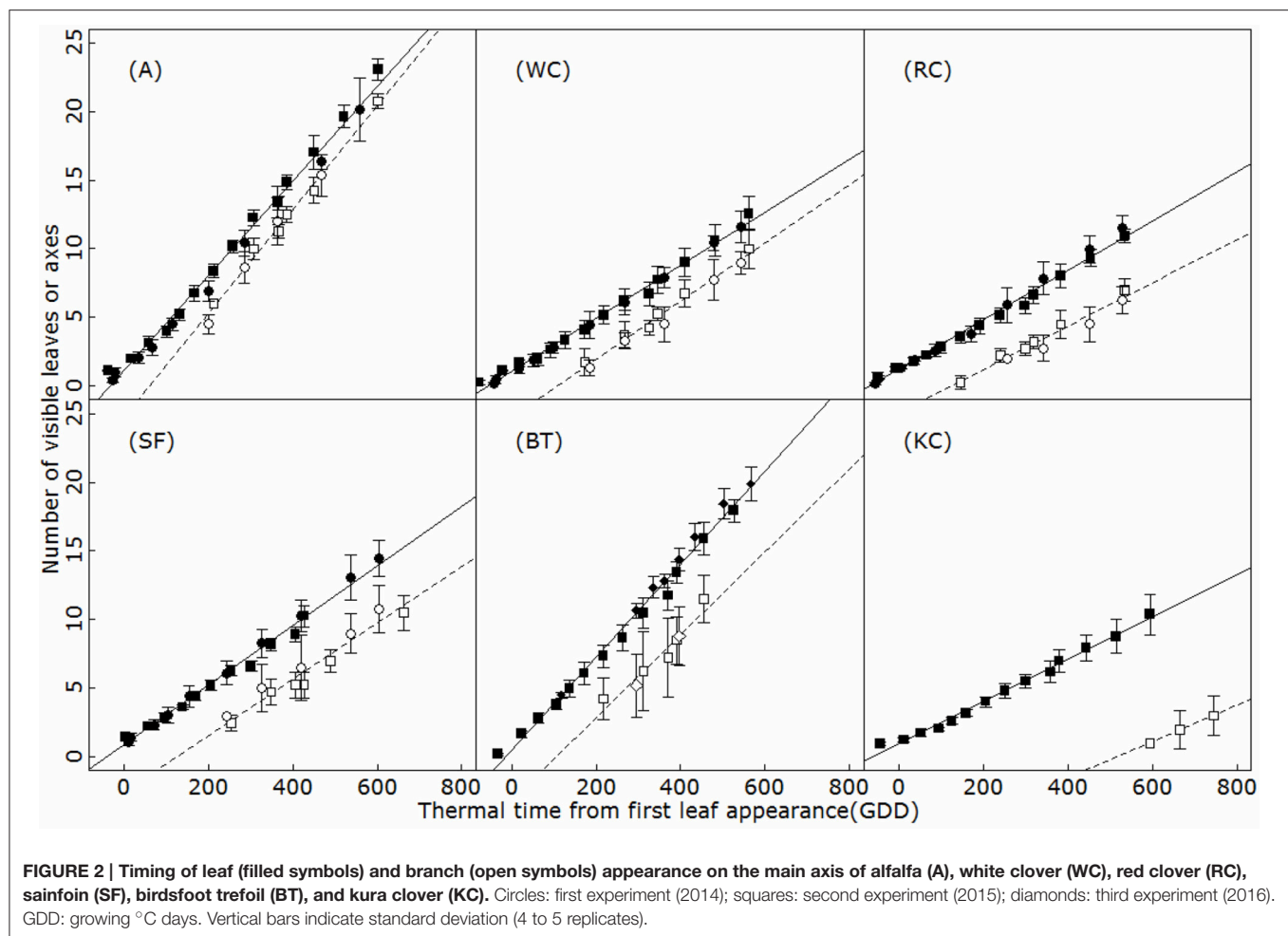
For each phytomer, branching probability was calculated at a given date as the ratio between the number of branches with an outburst at this position and the total number of plants in the treatment. The relationship between branching probability and phytomer position was characterized on the main and primary axes using a logistic function similar to Equation 3.

Significant differences between the means of plant traits were tested by performing analyses of variance ("aov" procedure). Analyses of covariance (ANCOVA, *lm* procedure) were used to test simultaneously for the effects of continuous and categorical variables and to compare the slopes and intercepts of linear relationships. Principal component analysis (PCA) was performed to assess the relationships between shoot morphogenetic parameters using the *ade4* package (13 parameters, 50 individuals).

RESULTS

Organogenesis on the Main Axis

New phytomers appeared on the main axis at a constant rate during the vegetative phase, resulting in a linear relationship between the total number of phytomers on an axis and thermal time in all the species studied ($R^2 > 0.95$; **Figure 2**). The leaf appearance rate differed markedly between species (ANCOVA, $p < 0.001$) and was conserved between experimental years.



Phyllochrons ranged from 32.6°C day in birdsfoot trefoil to 64.2°C day in Kura clover, with a stable pecking order (BT-A < SA < WC < RC < KC).

Branching

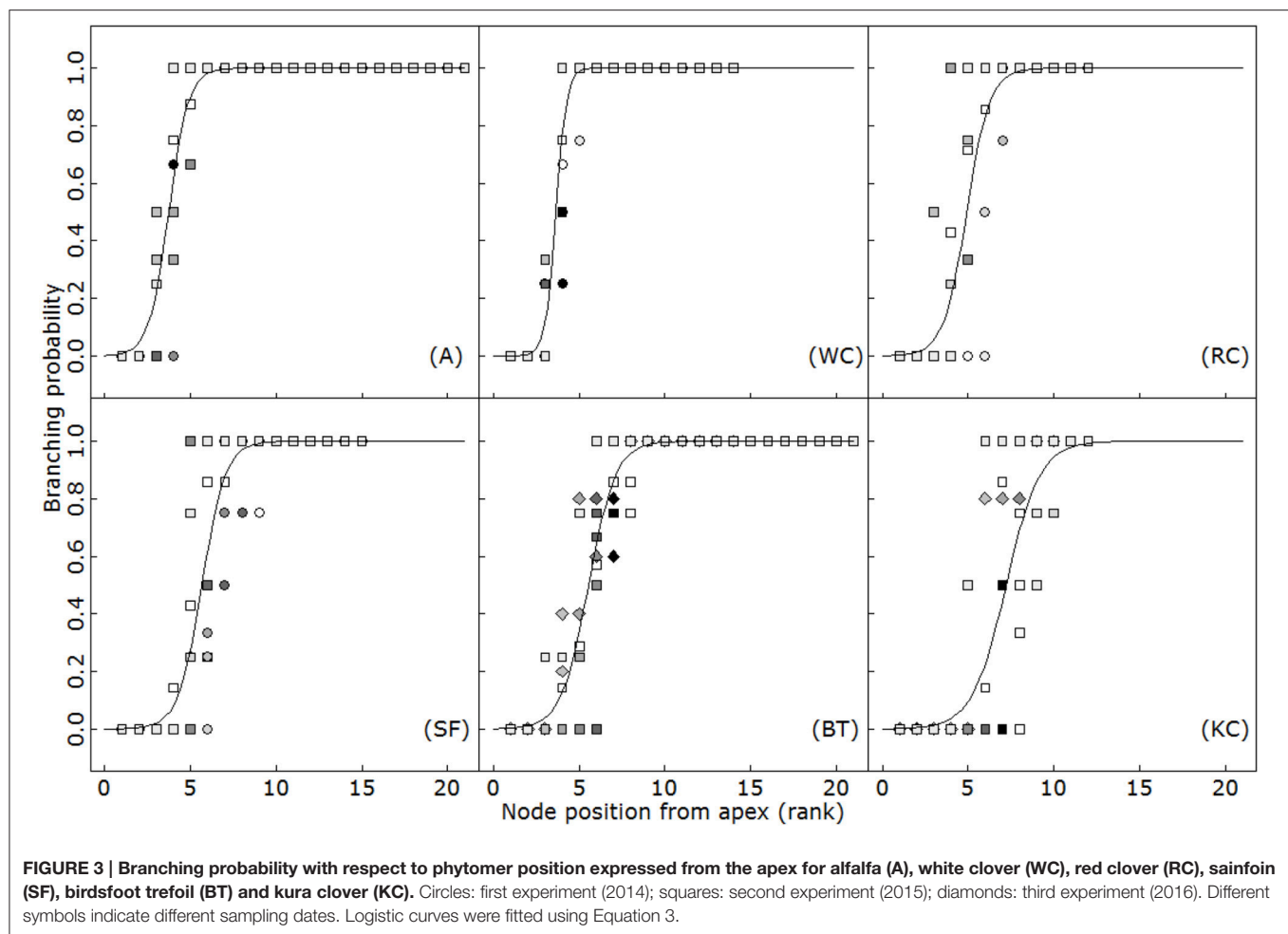
The timing of budburst of axillary branches on the main axis appeared to be linearly related to thermal time ($R^2 > 0.79$; **Figure 2**). A short lag period was observed systematically between the appearance of a phytomer and axillary budburst, so that phytomer appearance and the burst of the corresponding axillary bud was always separated by approximately two (A) to 8 (KC) phyllochrons, depending on the species. This delay of sylleptic branching remained constant over the period of observation in all the species, at all the positions (ANCOVA, $p > 0.77$). Accordingly, branching rates strongly differed between species and followed an inverse phyllochron order (BT-A > SA > WC > RC > KC). Furthermore, sylleptic branching appeared to be systematic on the main axis and primary axes (i.e., the branching probability ultimately reached 1 for all the phytomers after a certain delay; **Figure 3**), making it a deterministic process in isolated plants subjected to weak competition for light.

Organogenesis on the Primary and Secondary Axes

The development of primary (emerging from the collar zone) and secondary axes is further presented in **Figure 4**. As for the main axis, primary and secondary axes produced new phytomers at a constant rate in thermal time. Leaf appearance rates on the primary axes were similar to the main axis in A and BF, but differed in the other species (ANCOVA, $p < 0.01$). In white and red clovers, the primary axes developed more rapidly than the main axis, whereas the reverse was observed in sainfoin and Kura clover. Comparatively, the secondary axes displayed a much slower rate of development than primary axes (i.e., in A and BT, ANCOVA, $p < 0.001$). No significant effect of the topological position was found on the rate of development of primary and secondary axes in any of the six species (ANCOVA, $p > 0.20$).

Coordination of Organ Growth within a Phytomer

Once initiated, the different organs within a phytomer displayed highly conserved kinetics of expansion when expressed according to axis development (**Figure 5**). A strict scheduling organized the sequence of the onset of organ growth and the subsequent



relative expansion of the different organs. For all of them, the time sequence of organ elongation as a function of phyllochronic time was approximated correctly using a sigmoid function (Equation 3). No significant effects of phytomer rank or order were found with respect to growth delay and maximum relative elongation rate in any of the species (d and s parameters, ANOVA, $p > 0.51$), suggesting a coordination of organ growth was conserved irrespective of the phytomer within each species.

On the other hand, the different species displayed very dissimilar organ growth coordination patterns. The differences were particularly marked concerning the maximum elongation rates and the duration of phytomer expansion (ANOVA, $p < 0.001$). Some species, such as A and BT, typically presented a slow expansion of organs relative to axis development, and phytomer growth lasting for five phyllochrons. By contrast, white clover displayed rapid expansion and a total duration limited to 2.5 phyllochrons. The species also differed in the timing and relative order of organ growth. In most cases, leaf elements elongated first (leaflets > (stipules) > petiole), followed by internode elongation. However, in white clover, the order was reversed and internodes were first to complete their elongation.

Organ Dimensions at Maturity

For each species, the size ultimately attained by individual organs at maturity depended on the phytomer position. **Figure 6** presents the changes in relative organ dimensions along the primary axes. Irrespective of species and organ type, typical vegetative shoot patterns displayed profiles that first increased in size in line with phytomer ranks, and then stabilized at a plateau value, or even decreased. Interestingly, the relative profiles remained unchanged between the experimental years and were characteristic of isolated plants in a given species. Rank by rank comparisons of relative dimensions yielded identical mean values in 50 out of 53 cases for leaves, and 57 out of 58 cases for petioles/internodes (Student test, $p > 0.5$). For each organ type, the species differed in terms of the position of the phytomer at which maximum organ size was achieved. Concerning individual leaf size for instance, up to 12 phytomers were produced before reaching the peak in SF, but only 7–8 were necessary in WC or A.

As for the actual size of organs, maximum leaf area and petiole length differed in all the species, but the ranking between species remained unaffected by the experimental year (**Table 1**). Significant differences were observed between years only for the maximum internode length in some of the species. This trait was

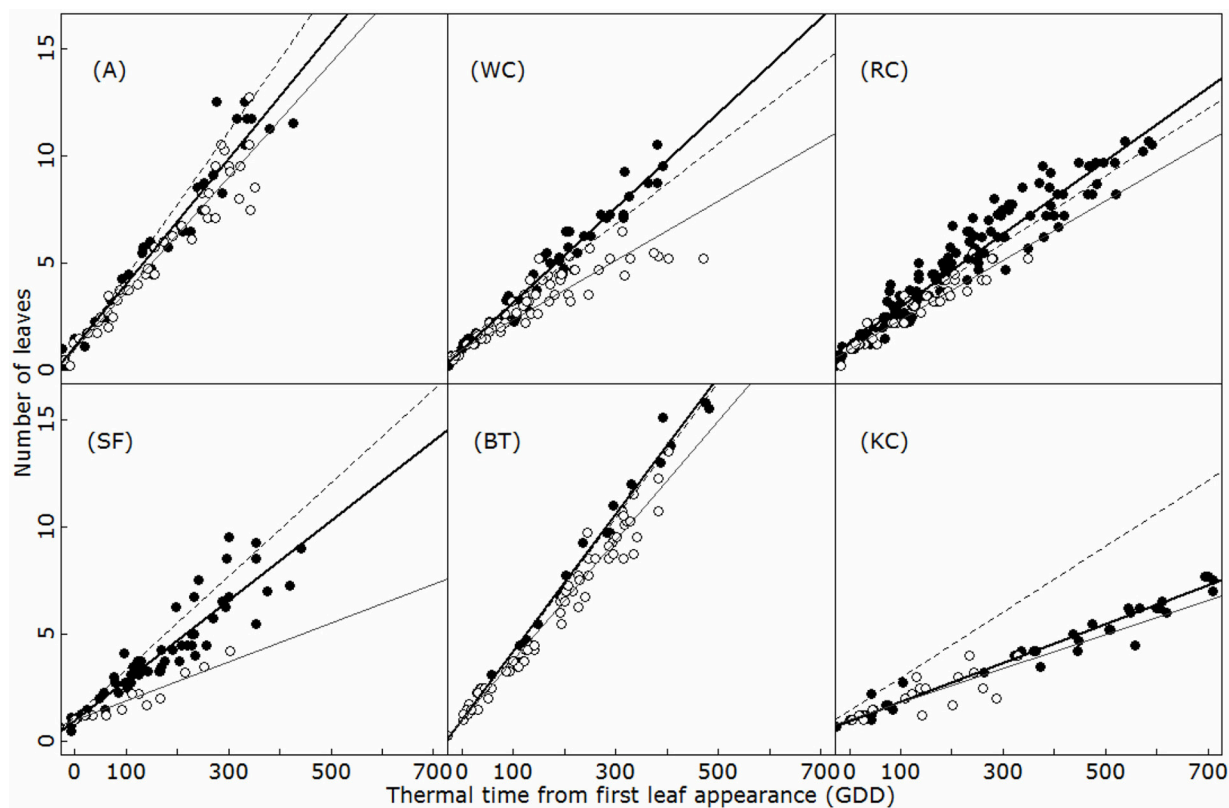


FIGURE 4 | Leaf appearance on the primary (filled circles) and secondary (open circles) axes for alfalfa (A), white clover (WC), red clover (RC), sainfoin (SF), birdsfoot trefoil (BT), and kura clover (KC). Dashed lines indicate the leaf appearance rate on the main axis (Figure 3). Bold and black lines indicate the linear regressions for the primary and secondary axes, respectively. Only axes with at least 3 unfolded leaves were plotted. GDD: growing °C days.

slightly smaller during the second experiment in white clover and alfalfa, but not in red clover and sainfoin.

Leaf Allometry

Leaf size and shape varied considerably between species and phytomers. Except in sainfoin, the leaves were all trifoliate from the second leaf on. Sainfoin presents compound leaves, with a number of leaflets that can increase up to 25 (Table 1). Despite the variability in their shapes, leaves from all the species complied with a single allometric relationship ($r^2 = 0.96$; Supplementary Figure S2) linking leaf area (LA) with central leaflet length (L), central leaflet width (l) and the number of leaflets (n):

$$LA = 0.694 \times L \times l \times n \quad (5)$$

Relationships between Shoot Morphogenetic Traits

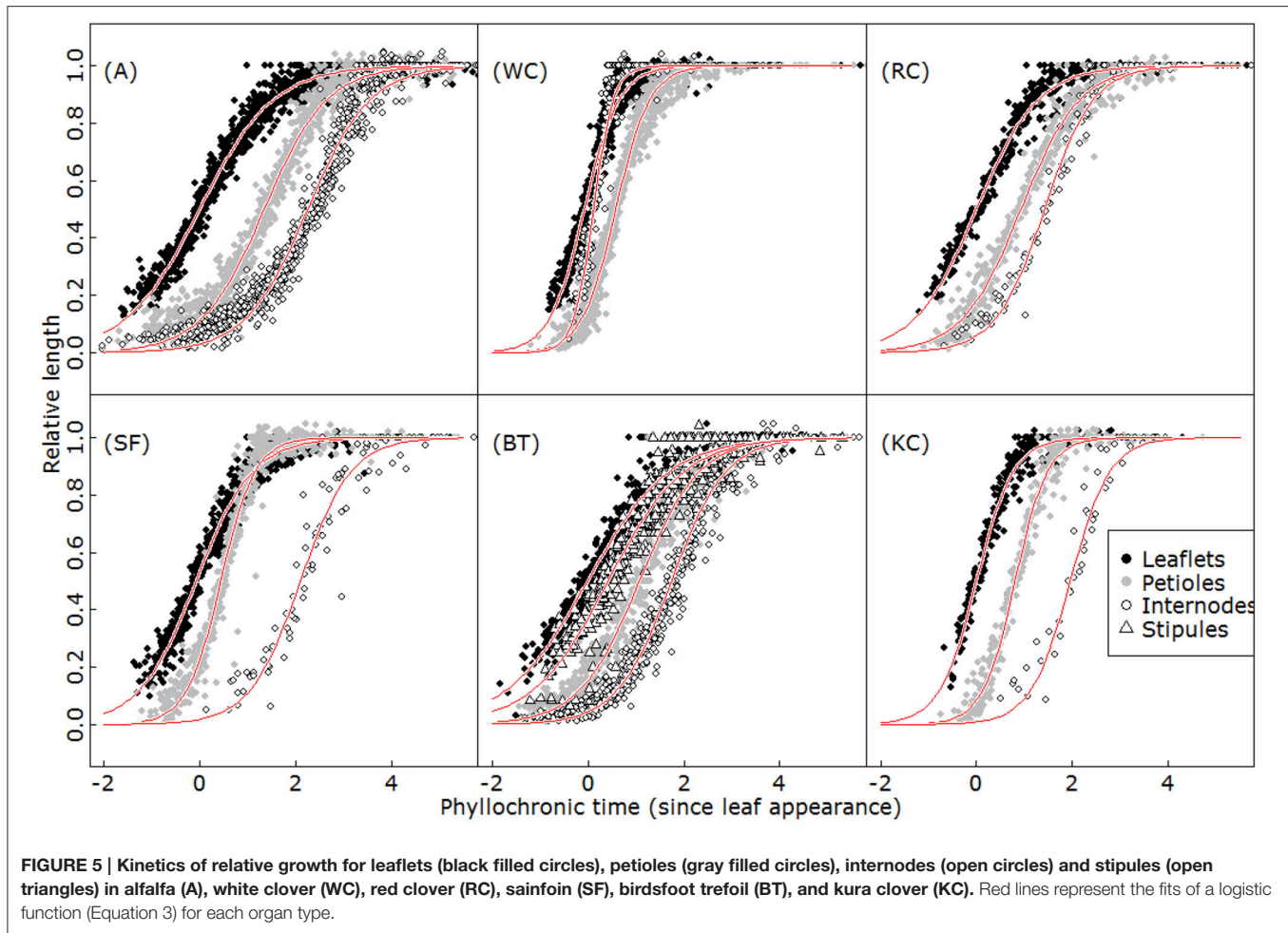
The possibility of defining sets of trait values occurring concomitantly was assessed by performing PCA on the dataset defined by the major morphogenetic traits characterized during the two experiments (Table 2). The first component of this PCA (Figure 7A), which explained almost half of total variance (47.8%), was mainly determined by leaf growth traits (MAX_{pet} , MAX_{lf} , duration of leaf growth) and by the organogenesis

of the main and primary axes (Phy0, Br1). It expressed an antagonism between the rate of phytomer production and the size and duration of expansion of leaf elements. Component 2, on the other hand, was mainly correlated to traits controlling the kinetics of internode expansion ($t50_{in}$) and to the development of secondary branches (ram_dist, Phy2). The growth of internodes was the most able in discriminating the species in terms of their growth coordination calendar. Petiole and leaflet expansions were tightly related in all the species, but internodes could expand either before or after leaf elements. In the plane containing component 1 and component 3 (not shown), it was seen that the third component was mainly correlated to the maximum length of internodes (MAX_{in}). Phy2 was also positively correlated with this third component, showing that species with the longest nodes also displayed a more vigorous development of secondary branches.

DISCUSSION

Potential Shoot Morphogenesis Followed a Set of Deterministic Rules Common to the Six Legume Species

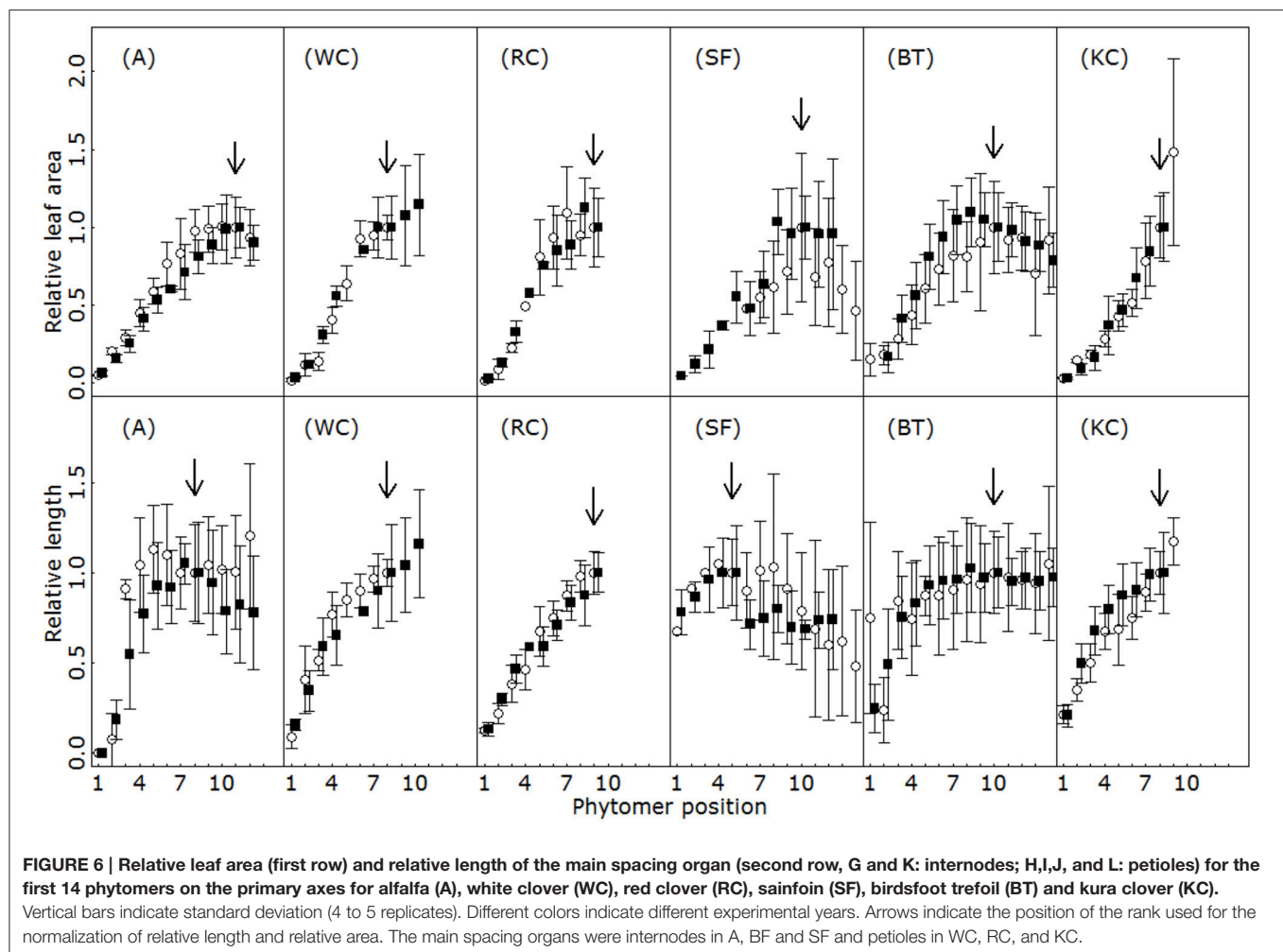
The morphogenesis of shoots can lead to highly differentiated plant architectures in perennial herbaceous legumes (Forde



et al., 1989; Thomas, 2003). However, as species differ in their branching complexity and in the size, position and shape of their shoot organs, our results highlighted the fact that they also share a number of determinants regarding the organogenesis and growth of phytomers, the building blocks of plant architecture. The striking differences between the species we studied emerged within the framework of a common and quite delimited pattern of vegetative growth and development. This framework involved: (i) a deterministic branching pattern and a regular development of shoot axes, (ii) a coordination of organ growth at the phytomer level, and (iii) a conserved allometry of leaf shapes.

Concerning shoot organogenesis, the existence of a generalized pattern of development was supported by the fact that all the species complied with the proposed classification of shoot axes in three categories. The main axis displayed developmental characteristics distinct from primary axes (that subsequently emerged close to the plant collar) and secondary branches (emerging from axillary buds out of the collar zone), each type presenting identical phyllochrons during the different experiments and for axes at different topological positions. In two of the species (namely A and BT), the main axis and primary axes presented similar characteristics, making the

classification even potentially simpler in legumes where the main axis can elongate. Given the important differences existing among dicots in terms of branching behavior and apical dominance (McSteen and Leyser, 2005), such regularity in the organogenetic process was not necessarily to be expected. Sylleptic shoot branching often occurs as a seemingly stochastic process, under the dependence of internal and environmental regulatory signals (Génard et al., 1994; Seleznyova et al., 2002; Rameau et al., 2015). Complex trophic and hormonal interplays can result in branches of the same order expressing very different characteristics depending on their position in the branching system (Lebon et al., 2004; Moreau et al., 2007), or presenting properties that change over time (e.g., increasing phyllochron, Barillot et al., 2012). Overall, the constant phyllochrons and branching delays we reported in the species we studied appeared to constitute a very simple way to characterize potential organogenesis. These characteristics may hold true mainly because we focused our attention on the period of vegetative development, during which little competition from other plant parts occurs. However, these observations were consistent with previous reports regarding herbaceous legumes (e.g., in alfalfa, Baldissera et al., 2014), and will make



the characterization of new species and genotypes easier in this group of species.

The temporal coordination of organ growth also appeared to be highly conserved, irrespective of phytomer position, in all the species. Such stable calendars of expansion had previously been reported for leaves and internodes on the primary axis of different dicotyledonous species (Granier et al., 2002; Demotes-Mainard et al., 2013). Interestingly, our results suggest that this could be extended to axes of different types once time is normalized by the rate of leaf appearance on each axis (i.e., phyllochronic time). Such an approach is new for dicots but had been used successfully to account for the flexible growth pattern of grass leaves, constantly adapting to the timing of leaf appearance (Fournier et al., 2005; Zhu et al., 2014). It could probably be applied to account for the relationship between the rate of phytomer production and the timing and duration of organ growth within a phytomer on a broad range of species.

The allometric relationship found between leaf area and the product of leaf length and leaf width is currently applied in numerous species (Schwarz and Kläring, 2001; Antunes et al., 2008; Baldissera et al., 2014). Accounting for the number

of leaflets was sufficient to encompass the different species within the same relationship in our study (Equation 5). This indicates that a common shape coefficient could be applied to expanded leaflets from the different legume species (Prévot et al., 1991). However, considering both leaf length and leaf width was a necessity to establish this common relationship, because some dissociation exists between the longitudinal and lateral expansion of leaf laminae (e.g., the former being mediated by brassinosteroids without any effect on the latter; Nakaya et al., 2002), resulting in length to width ratios that vary as a function of phytomer position and species.

On the other hand, no clear pattern governing organ dimension at maturity, and which could encompass all the species, emerged at the axis level. Typical leaf and internode length profiles reached a maximum value at an intermediate position along the axis, a situation that is relatively ubiquitous in vascular plants (Allsopp, 1967; Villani and Demason, 2000). These positions were similar between experiments in a given species, but they changed dramatically depending on organ type and between species. Furthermore, the maximum dimension attained by internodes changed between two of the experiments.

TABLE 1 | Maximum organ dimensions observed during experiments 1 and 2 for alfalfa (A), white clover (WC), red clover (RC), sainfoin (SA), birdsfoot trefoil (BT), and kura clover (KC).

Organ	A			WC			RC			SF			BT			KC			ANOVA						
	2014		2015	2014		2015	2014		2015	2014		2015	2014		2015	2015		Species		Year				Species x Year	
																		F		Pr		F		Pr	
																		F		Pr		F		Pr	
Petiole length	61.5 ± 4.5	72.5 ± 6.1	261 ± 31	259 ± 28	290 ± 35	255 ± 34	101 ± 34	77.5 ± 12.0	10.6 ± 0.95	208 ± 28	137.21	<0.001	1.693	0.21	1.21	0.33									
Internode length	85.0 ± 3.92	56.8 ± 8.46	49.0 ± 7.53	38.1 ± 4.67	87.8 ± 4.5	84.2 ± 30.4	68.0 ± 16.3	74.1 ± 1.21	24.2 ± 8.42	62.7 ± 9.45	15.90	<0.001	6.68	<0.05	2.70	<0.05									
Number of leaflets	3	3	3	3	3	3	15 ± 3.5	16.5 ± 7.9	3	3	110.50	<0.001	0.04	0.83	0.04	0.99									
Leaf area	33.1 ± 8.3	28.5 ± 1.8	46 ± 5.0	46.7 ± 5.9	64.3 ± 11.0	62.2 ± 8.3	73.9 ± 32.2	58.3 ± 12.7	8.9 ± 2.2	37.7 ± 15.6	11.35	<0.001	1.23	0.28	0.53	0.66									

Average ± standard deviation (length in mm, area in cm²). Only the four species measured during the two experiments were included in the ANOVA.

TABLE 2 | Variables used in the Principal Component Analysis presented in Figure 7.

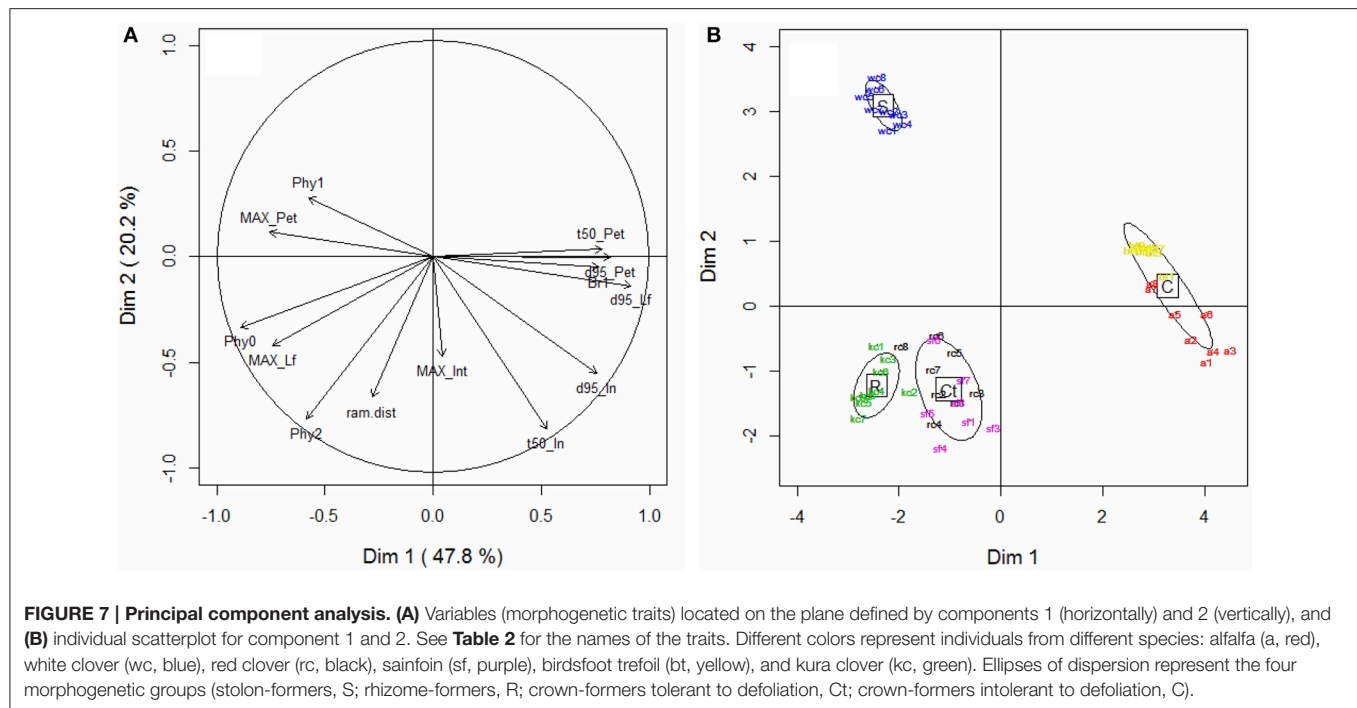
Abbreviation	Variable	Unit
Phy0	Phyllochron of the main axis	°Cd. phytomer ⁻¹
Br1	Primary axis production rate	Phyllochron.axis ⁻¹
Phy1	Phyllochron of primary axes relative to the main axis	Phytomer.phytomer ⁻¹
Ram.dist	Delay between phytomer production and axillary budburst	Phyllochron
d95_in	Internode expansion duration	Phyllochron
d95_pet	Petiole expansion duration	Phyllochron
d95_lf	Leaflet expansion duration	Phyllochron
t50_pet	Delay between leaflet and petiole expansion	Phyllochron
t50_in	Delay between leaflet and internode expansion	Phyllochron
MAX _{pet}	Maximal petiole length at maturity	cm
MAX _{in}	Maximal internode length at maturity	cm
MAX _{lf}	Maximal leaf area at maturity	cm ²

As water and nutrients were supplied without restriction, these differences could have resulted from the slightly different light regimes prevailing in the different years, or could have been caused by greater evaporative demand affecting organ expansion in 2015 (Supplementary Table S1, Tardieu et al., 2005).

Differences in Trait Values Determined the Contrasting Morphogenetic Strategies

Despite this common developmental pattern, the species sampled did produce contrasting shoot architectures. Even without considering geometric features (such as leaf angles, shoot bearing, etc.), the values for component traits of shoot morphogenesis and which accounted for phytomer production (phyllochrons, delay of branching) and organ growth (delays of expansion, relative expansion rate, maximum organ dimension) explained the emergence of four morphogenetic groups (Figure 7B) that partially matched those identified by Thomas (2003). Only kura clover, which is in fact a rhizome-forming species, was closely related with two crown-forming species producing a short main axis (RC and SF). As previously shown (Genrich et al., 1998; Black et al., 2006), kura clover was very slow to develop and did not express its rhizomatous growth habit during the period studied. The primary axes characterized from rhizome cuttings in the third experiments did not differ from those emerging from collar in the second experiment. The developmental pattern of KC shoots thus appeared very similar to red clover in our conditions, explaining the close classification of the two species.

The main morphogenetic traits associated with this classification were distinguished along two PCA axes. A first dimension represented the strategies of leaf area production, reliant on either the production of numerous small leaves or the expansion of a fewer large leaves. Such a trade-off between growth and development is common during morphogenesis and has, for instance, been reported in grass tillers (leaf length



vs. phyllochron and tillering rate; Gautier et al., 1999; Nelson, 2000), and in roots (root elongation rate vs. branching ability: Pagès, 2014). In forage legumes, this could be a key element in controlling the rate of regrowth in a mixture after defoliation (e.g., in white clover and alfalfa: Davies, 2001; Annicchiarico, 2003; Annicchiarico et al., 2015). A second dimension mainly distinguished species in terms of their growth coordination patterns, promoting either early internode growth or the early growth of leaf elements (including petioles) as a primary option to colonize space and expand vegetative structures. All species but white clover favored the growth of leaf elements, which is consistent with a pre-emptive strategy for light acquisition that could be expected in crown-formers with limited prospecting ability. Under such local competition, rapid plant leaf area development indeed appears to be the most successful trait to capture light (Louarn et al., 2012). On the other hand, white clover develops primary axes by first sensing its environment through internode elongation. This particular trait could be an adaptation to achieve horizontal prospection for light gaps, as white clover tends to be a light foraging species rather than a competitive one (de Kroons and Hutchings, 1995). Similar growth coordination patterns are observed in lianas (e.g., grapevine, Louarn, 2005) and in invasive species sprouting through long cane emission (e.g., *rubus* sp., Amor, 1974), thus making such an hypothesis plausible.

A degree of redundancy between several of the morphogenetic traits was apparent (e.g., high correlations between petiole and leaflet size, or between the size of these organs and their duration of expansion), which suggests opportunities to simplify the characterization of morphogenetic strategies. Overall, however, the combined values of these traits enabled the discrimination

and classification of the species as a function of their known ecological behaviors.

Interests and Limitations of Such a Framework for Plant Phenotyping and Modeling

Identifying a robust framework that enables the description, analysis and prediction of plant morphogenesis is central to developing efficient phenotyping approaches for breeding (Granier et al., 2002; Moreau et al., 2006) and the diagnosis of plant stress (e.g., Pellegrino et al., 2005). The framework for potential morphogenesis we have discussed above represents a first step toward achieving both of these goals with respect to perennial herbaceous legumes, many of which have been little characterized to date. Setting up such a potential framework in absence of competition is particularly important to identify ontogenic patterns of development and to disentangle the contradictory effects of neighbors on plant morphology and resource acquisition (Lemaire and Millard, 1999).

Clearly, our approach is currently limited to the vegetative period of development, and the interactions which may take place during flowering or the development of reproductive organs at later stages are not taken into account. In most forage species, regular harvests of shoot biomass over a season involve several cuts of primary shoot axes, and a succession of vegetative development from crown and the remaining lateral buds, thus making particularly relevant the vegetative framework presented. However, in prostrate species such as white clover, or in those with a terminal flowering (Thomas, 2003), interactions with the reproductive cycle will occur at some point, and integrating a reproductive dimension in the framework would most likely

improve its robustness (e.g., Moreau et al., 2007). Similarly, the development of nodal roots has been shown to interact with shoot organogenesis in aging clover plants (Thomas et al., 2014), and a whole plant appraisal of morphogenesis, considering both shoot and root morphogenetic traits (Pagès, 2014) would deserve exploration. In any case, the set of deterministic rules identified in this study could serve as an initial benchmark for the future development of a more comprehensive framework. As a potential morphogenetic model, it could also serve as a baseline to analyse the responses of vegetative development to environmental stresses (Belaygue et al., 1996; Lebon et al., 2006; Baldissera et al., 2014), and to compare on a quantitative basis a broad range of species and genotypes with respect to their morphogenetic strategies (Louarn et al., 2007).

At present, the framework provides a generic approach to break down the component traits of competitive ability aboveground (i.e., the components of leaf area expansion and height acquisition: Barillot et al., 2011; Louarn et al., 2012). This is of particular value because: (i) the genetic component of competitive ability has been shown to be more tightly related to component traits of morphogenesis than to integrated traits (Annicchiarico et al., 1999), (ii) such a framework could help to highlight specific combinations of traits associated with a particular morphogenetic group, even in species for which the amount of background studies is limited, and (iii) the knowledge of component traits associated with competitive ability is expected to increase the efficiency of pure stand selection that targets mixed stand conditions, thus offering opportunities to reduce costs during the selection process for multi-species grasslands (Annicchiarico, 2003). Finally,

such a quantitative approach could also readily be integrated as a component in mixed grassland models (Soussana et al., 2012), so as to enable better consideration of the diversity of forms that legume components could accommodate, and to improve the prediction of their persistence in mixtures.

AUTHOR CONTRIBUTIONS

LF and GL designed the experiments, conducted measurements and performed data analyses. IL and GL rose funding for these research. All of the authors contributed to writing the manuscript.

ACKNOWLEDGMENTS

This study was supported by the Agence Nationale de la Recherche (PRAISE project, ANR-13-BIOADAPT-0015), the Poitou Charentes Regional Council (PhD fellowship for LF) and INRA's Environment and Agronomy Division (Transfert-N Project, PhD fellowship for LF). The authors would like to thank M. Demon, A. Eprinchard, C. Perrot, E. Roy, E. Rivault, A. Philipponneau, and J. P. Terrasson for their assistance with the experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00405/full#supplementary-material>

REFERENCES

- Allsopp, A. (1967). Heteroblastic development in vascular plants. *Adv. Morphog.* 6, 127–171. doi: 10.1016/b978-1-4831-9953-5.50008-1
- Amor, R. L. (1974). Ecology and control of blackberry (*Rubus fruticosus* L. agg.) II. Reproduction. *Weed Res.* 14, 231–238. doi: 10.1111/j.1365-3180.1974.tb01047.x
- Annicchiarico, P. (2003). Breeding white clover for increased ability to compete with associated grasses. *J. Agric. Sci.* 140, 255–266. doi: 10.1017/S0021859603003198
- Annicchiarico, P., Barrett, B., Brummer, E. C., Julier, B., and Marshall, A. H. (2015). Achievements and challenges in improving temperate perennial forage legumes. *CRC Crit. Rev. Plant Sci.* 34, 327–380. doi: 10.1080/07352689.2014.898462
- Annicchiarico, P., Piano, E., and Rhodes, I. (1999). Heritability of, and genetic correlations among, forage and seed yield traits in ladino white clover. *Plant Breed.* 118, 341–346. doi: 10.1046/j.1439-0523.1999.00387.x
- Antunes, W. C., Pompelli, M. F., Carretero, D. M., and DaMatta, F. M. (2008). Allometric models for non-destructive leaf area estimation in coffee (*Coffea arabica* and *Coffea canephora*). *Ann. Appl. Biol.* 153, 33–40. doi: 10.1111/j.1744-7348.2008.00235.x
- Baldissera, T. C., Frak, E., de Faccio Carvalho, P. C., and Louarn, G. (2014). Plant development controls leaf area expansion in alfalfa plants competing for light. *Ann. Bot.* 113, 145–157. doi: 10.1093/aob/mct251
- Barillot, R., Chevalier, V., Combes, D., and Louarn, G. (2012). “Variability of pea morphogenesis. An exploratory study based on six pea genotypes with contrasting architectures,” in *Fourth International Symposium on Plant Growth Modelling, Simulation, Visualization and Applications (PMA12)*, eds M. Kang, Y. Dumont, and Y. Guo (Shanghai: IEEE Press), 43–49.
- Barillot, R., Escobar-Gutiérrez, A. J., Fournier, C., Huynh, P., and Combes, D. (2014). Assessing the effects of architectural variations on light partitioning within virtual wheat-pea mixtures. *Ann. Bot.* 114, 725–737. doi: 10.1093/aob/mcu099
- Barillot, R., Louarn, G., Escobar-Gutiérrez, A. J., Huynh, P., and Combes, D. (2011). How good is the turbid medium-based approach for accounting for light partitioning in contrasted grass-legume intercropping systems? *Ann. Bot.* 108, 1013–1024. doi: 10.1093/aob/mcr199
- Barre, P., Turner, L. B., and Escobar-Gutiérrez, A. J. (2015). Leaf length variation in perennial forage grasses. *Agriculture* 5, 682–696. doi: 10.3390/agriculture5030682
- Belaygue, C., Wery, J., Cowan, A., and Tardieu, F. (1996). Contribution of leaf expansion, rate of leaf appearance, and stolon branching to growth of plant leaf area under water deficit in white clover. *Crop Sci.* 36, 1240–1246. doi: 10.2135/cropsci1996.0011183x003600050028x
- Beuselinck, P. R., Bouton, J. H., Lamp, W. O., Matches, A. G., McCaslin, M. H., Nelson, C. J., et al. (1994). Improving legume persistence in forage crop systems. *J. Prod. Agric.* 7, 311–322. doi: 10.2134/jpa1994.0311
- Black, A. D., Moot, D. J., and Lucas, R. J. (2006). Development and growth characteristics of Caucasian and white clover seedlings, compared with perennial ryegrass. *Grass Forage Sci.* 61, 442–453. doi: 10.1111/j.1365-2494.2006.00553.x
- Briske, D. D. (1996). “Strategies of plant survival in grazed systems: a functional interpretation,” in *The Ecology and Management of Grazing Systems*, eds J. Hodgson and A. W. Illius (Wallingford: CAB International), 37–67.
- Caldwell, M. M. (1987). “Plant architecture and resource competition,” in *Potentials and Limitations of Ecosystem Analysis - Ecological Studies*, 61, eds E. D. Schulze and H. Zwölfer (Berlin-Heidelberg: Springer-Verlag), 164–179.

- Costes, E., and Guédon, Y. (2002). Modelling branching patterns on 1-year-old trunks of six apple cultivars. *Ann. Bot.* 89, 513–524. doi: 10.1093/aob/mcf078
- Davies, A. (2001). “Competition between grasses and legumes in established pastures,” in *Competition and Succession in Pastures*, eds P. G. Tow and A. Lazenby (Oxon: CABI Publishing), 63–83.
- de Kroons, H., and Hutchings, M. J. (1995). Morphological plasticity in clonal plants: the foraging concept reconsidered. *J. Ecol.* 83, 143–152. doi: 10.2307/2261158
- Demotes-Mainard, S., Bertheloot, J., Boumaza, R., Huché-Thélier, L., Guéritaine, G., Guérin, V., et al. (2013). Rose bush leaf and internode expansion dynamics: analysis and development of a model capturing interplant variability. *Front. Plant Sci.* 4:418. doi: 10.3389/fpls.2013.00418
- Forde, M. B., Hay, M. J. M., and Brock, J. L. (1989). “Development and growth characteristics of temperate perennial legumes,” in *Persistence of Forage Legume*, eds G. C. Marten, A. G. Matches, R. F. Barnes, R. W. Brougham, R. J. Clements, and G. W. Sheath (Madison, WI: American Society of Agronomy), 91–110.
- Fort, F., Cruz, P., Catrice, O., Delbrut, A., Luzarreta, M., Stroia, C., et al. (2015). Root functional trait syndromes and plasticity drive the ability of grassland Fabaceae to tolerate water and phosphorus shortage. *Environ. Exp. Bot.* 110, 62–72. doi: 10.1016/j.envexpbot.2014.09.007
- Fournier, C., Durand, J. L., Ljutovac, S., Schäufele, R., Gastal, F., and Andrieu, B. (2005). A functional-structural model of elongation of the grass leaf and its relationships with the phyllochron. *New Phytol.* 166, 881–894. doi: 10.1111/j.1469-8137.2005.01371.x
- Gaba, S., Lescouret, F., Boudsocq, S., Enjalbert, J., Hinsinger, P., Journet, E. P., Navas, M. L., et al. (2015). Multiple cropping systems as drivers for providing multiple ecosystem services: from concepts to design. *Agron. Sustain. Dev.* 35, 607–623. doi: 10.1007/s13593-014-0272-z
- Gastal, F., and Saugier, B. (1986). Nitrogen nutrition and growth of tall fescue. 2. Nitrogen uptake and partitioning among plant organs. *Agronomie* 6, 363–370. doi: 10.1051/agro:19860406
- Gautier, H., Mèch, R., Prusinkiewicz, P., and Varlet-Grancher, C. (2000). 3D Architectural modelling of aerial photomorphogenesis in white clover (*Trifolium repens* L.) using L-systems. *Ann. Bot.* 85, 359–370. doi: 10.1006/anbo.1999.1069
- Gautier, H., Varlet-Grancher, C., and Hazard, L. (1999). Tillering responses to the light environment and to defoliation in populations of perennial ryegrass (*Lolium perenne* L.) selected for contrasting leaf length. *Ann. Bot.* 83, 423–429. doi: 10.1006/anbo.1998.0840
- Génard, M., Pagès, L., and Kervella, J. (1994). Relationship between sylleptic branching and components of parent shoot development in the peach tree. *Ann. Bot.* 74, 465–470. doi: 10.1006/anbo.1994.1142
- Genrich, K. C., Sheaffer, C. C., and Ehlike, N. J. (1998). Kura clover growth and development during the seeding year. *Crop Sci.* 38, 735–741. doi: 10.2135/cropsci1998.0011183x003800030020x
- Gosse, G., Lemaire, G., Chartier, M., and Balfourier, F. (1988). Structure of a lucerne population (*Medicago sativa* L.) and dynamics of stem competition for light during regrowth. *J. Appl. Ecol.* 25, 609–617. doi: 10.2307/2403848
- Granier, C., Massonnet, C., Turc, O., Muller, B., Chenu, K., and Tardieu, F. (2002). Individual leaf development in *Arabidopsis thaliana*: a stable thermal-time-based programme. *Ann. Bot.* 89, 595–604. doi: 10.1093/aob/mcf085
- Jensen, E. S., Peoples, M. B., Boddey, R. M., Gresshoff, P. M., Hauggaard-Nielsen, H., Alves, B. J. R., et al. (2012). Legumes for mitigation of climate change and the provision of feedstock for biofuels and biorefineries. A review. *Agron. Sustain. Dev.* 32, 329–364. doi: 10.1007/s13593-011-0056-7
- Kraft, N. J., Godoy, O., and Levine, J. M. (2015). Plant functional traits and the multidimensional nature of species coexistence. *Proc. Natl. Acad. Sci. U.S.A.* 112, 797–802. doi: 10.1073/pnas.1413650112
- Lafarge, M., and Durand, J. L. (2013). *Comment L'herbe Pousse: Développement Végétatif, Structures Clonales et Spatiales des Graminées*. Versailles: Quae Editions.
- Lebon, E., Pellegrino, A., Louarn, G., and Lecoeur, J. (2006). Branch development controls leaf area dynamics in grapevine (*Vitis vinifera*) growing in drying soil. *Ann. Bot.* 98, 175–185. doi: 10.1093/aob/mcl085
- Lebon, E., Pellegrino, A., Tardieu, F., and Lecoeur, J. (2004). Shoot development in grapevine (*Vitis vinifera*) is affected by the modular branching pattern of the stem and intra- and inter-shoot trophic competition. *Ann. Bot.* 93, 263–274. doi: 10.1093/aob/mch038
- Lecoeur, J. (2010). “Vegetative development,” in *Physiology of the Pea Crop*, eds N. Munier-Jolain, V. Biarnès, I. Chaillat, J. Lecoeur, and M. H. Jeuffroy (Boca Raton, FL: CRC Press), 3–43.
- Lemaire, G., and Millard, P. (1999). An ecophysiological approach to modelling resource fluxes in competing plants. *J. Exp. Bot.* 50, 15–28. doi: 10.1093/jxb/50.330.15
- Lewis, G., Schrire, B., Mackinder, B., and Lock, M. (2005). *Legumes of the World*. Kew: Royal Botanic Gardens.
- Louarn, G. (2005). *Analyse et Modélisation de l'organogenèse et de l'architecture du Rameau de Vigne (Vitis vinifera L.)*. PhD thesis Montpellier: ENSA.
- Louarn, G., Andrieu, B., and Giauffret, C. (2010). A size-mediated effect can compensate for transient chilling stress affecting maize (*Zea mays*) leaf extension. *New Phytol.* 187, 106–118. doi: 10.1111/j.1469-8137.2010.03260.x
- Louarn, G., Da Silva, D., Godin, C., and Combes, D. (2012). Simple envelope-based reconstruction methods can infer light partitioning among individual plants in sparse and dense herbaceous canopies. *Agric. Forest Meteorol.* 166, 98–112. doi: 10.1016/j.agrformet.2012.06.014
- Louarn, G., Guedon, Y., Lecoeur, J., and Lebon, E. (2007). Quantitative analysis of the phenotypic variability of shoot architecture in two grapevine (*Vitis vinifera*) cultivars. *Ann. Bot.* 99, 425–437. doi: 10.1093/aob/mcl276
- LPWG. (2013). Global legume diversity assessment: concepts, key indicators, and strategies. *Taxon* 62, 249–266. doi: 10.12705/622.12
- Maamouri, A., Louarn, G., Gastal, F., Béguier, V., and Julier, B. (2015). Effects of lucerne genotype on morphology, biomass production and nitrogen content of lucerne and tall fescue in mixed pastures. *Crop Past. Sci.* 66, 192–204. doi: 10.1071/CP14164
- Maitre, J. P., Assemat, L., and Jacquard, P. (1985). Croissance du trèfle violet (*Trifolium pratense* L.) en association avec du ray-grass d'Italie (*Lolium multiflorum* Lam. ssp. italicum). I. – Description de l'organisation morphologique du trèfle. *Agronomie* 5, 251–60. doi: 10.1051/agro:19850307
- McSteen, P., and Leyser, O. (2005). Shoot branching. *Annu. Rev. Plant Biol.* 56, 353–374. doi: 10.1146/annurev.arplant.56.032604.144122
- Moreau, D., Salon, C., and Munier-Jolain, N. (2006). Using a standard framework for the phenotypic analysis of *Medicago Truncatula*: an effective method for characterizing the plant material used for functional genomics approaches. *Plant Cell Environ.* 29, 1087–1098. doi: 10.1111/j.1365-3040.2005.01483.x
- Moreau, D., Salon, C., and Munier-Jolain, N. (2007). A Model-based framework for the phenotypic characterization of the flowering of *Medicago Truncatula*. *Plant Cell Environ.* 30, 213–224. doi: 10.1111/j.1365-3040.2006.01620.x
- Mouliat, B., Edelin, C., Jeuffroy, M. H., Allirand, J. M., Loup, C., and Chartier, M. (2000). “Preliminary analysis of the architecture of cultivated herbaceous plants,” in *Fonctionnement des Peuplements Végétaux Sous Contraintes Environnementales*, eds P. Maillard and R. Bonhomme (Paris: INRA Editions), 149–184.
- Nakaya, M., Tsukaya, H., Murakami, N., and Kato, M. (2002). Brassinosteroids control the proliferation of leaf cells of *Arabidopsis Thaliana*. *Plant Cell Physiol.* 43, 239–244. doi: 10.1093/pcp/pcf024
- Nelson, C. J. (2000). “Shoot morphological plasticity of grasses: leaf growth vs. tillering,” in *Grassland Ecophysiology and Grazing Ecology*, eds G. Lemaire, J. Hodgson, A. D. Moraes, P. D. F. Carvalho, and C. Nabinger (New York, NY: CAB International Publishing), 101–126.
- Nyfelers, D., Huguenin-Elie, O., Suter, M., Frossard, E., and Lüscher, A. (2011). Grass-legume mixtures can yield more nitrogen than legume pure stands due to mutual stimulation of nitrogen uptake from symbiotic and non-symbiotic sources. *Agric. Ecosyst. Environ.* 140, 155–163. doi: 10.1016/j.agee.2010.11.022
- Owens, L. B., Edwards, W. M., and Van Keuren, R. W. (1994). Groundwater nitrate levels under fertilized grass and grass-legume pastures. *J. Environ. Qual.* 23:752. doi: 10.2134/jeq1994.00472425002300040020x
- Pagès, L. (2014). Branching patterns of root systems: quantitative analysis of the diversity among dicotyledonous species. *Ann. Bot.* 114, 591–598. doi: 10.1093/aob/mcu145
- Pellegrino, A., Lebon, E., Simmoneau, T., and Wery, J. (2005). Towards a simple indicator of water stress in grapevine (*Vitis vinifera* L.) based on the differential sensitivities of vegetative growth components. *Aust. J. Grape Wine Res.* 11, 306–315. doi: 10.1111/j.1755-0238.2005.tb00030.x

- Prévot, L., Ariès, F., and Monestiez, P. (1991). Modélisation de la structure géométrique du maïs. *Agronomie* 11, 491–503. doi: 10.1051/agro:19910606
- Rameau, C., Bertheloot, J., Leduc, N., Andrieu, B., Foucher, F., and Sakr, S. (2015). Multiple pathways regulate shoot branching. *Front. Plant Sci.* 5:741. doi: 10.3389/fpls.2014.00741
- R Development Core Team (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Ross, J. K. (1981). *The Radiation Regime and Architecture of Plant Stands*. Hague: Dr. W. Junk.
- Russelle, M. P., Lamb, J. F. S., Montgomery, B. R., Elsenheimer, D. W., Miller, B. S., and Vance, C. P. (2001). Alfalfa rapidly remediates excess inorganic N at a fertilizer spill site. *J. Environ. Qual.* 30, 30–36. doi: 10.2134/jeq2001.30130x
- Schwarz, D., and Kläring, H. P. (2001). Allometry to estimate leaf area of tomato. *J. Plant Nutr.* 24, 1291–1309. doi: 10.1081/pln-100106982
- Schwinning, S., and Parsons, A. J. (1996). “Interactions between grasses and legumes: understanding variability in species composition,” in *Legumes in Sustainable Farming Systems, Occasional Symposium - British Grassland Society No. 30*, ed D. Younie (Reading: British Grassland Society), 153–163.
- Scott, D., Hoglund, J. H., Crush, J. R., and Keoghan, J. M. (1989). “Environmental selection of legumes,” in *Persistence of Forage Legume*, eds G. C. Marten, A. G. Matches, R. F. Barnes, R. W. Brougham, R. J. Clements, and G. W. Sheath (Madison, WI: American Society of Agronomy), 205–214.
- Seleznayova, A. N., Thorp, T. G., Barnett, A. M., and Costes, E. (2002). Quantitative analysis of shoot development and branching patterns in *Actinidia*. *Ann. Bot.* 89, 471–482. doi: 10.1093/aob/mcf069
- Sheaffer, C. C. (1989). “Effect of competition on legume persistence,” in *Persistence of Forage Legume*, eds G. C. Marten, A. G. Matches, R. F. Barnes, R. W. Brougham, R. J. Clements, and G. W. Sheath (Madison, WI: American Society of Agronomy), 327–333.
- Simon, J. C., and Lemaire, G. (1987). Tillering and leaf area index in grasses in the vegetative phase. *Grass Forage Sci.* 42, 373–380. doi: 10.1111/j.1365-2494.1987.tb02127.x
- Skinner, R. H., and Nelson, C. J. (1994). Epidermal cell division and the coordination of leaf and tiller development. *Ann. Bot.* 74, 9–16. doi: 10.1006/anbo.1994.1088
- Smith, S. R. Jr., Bouton, J. H., Singh, A., and McCaughey, W. P. (2000). Development and evaluation of grazing-tolerant alfalfa cultivars: A review. *Can. J. Plant Sci.* 80, 503–512. doi: 10.4141/p99-048
- Soussana, J. F., Maire, V., Gross, N., Bachelet, B., Pagès, L., Martin, R., et al. (2012). Gemini: a grassland model simulating the role of plant traits for community dynamics and ecosystem functioning. Parameterization and evaluation. *Ecol. Model.* 231, 134–145. doi: 10.1016/j.ecolmodel.2012.02.002
- Suter, M., Connolly, J., Finn, J. A., Loges, R., Kirwan, L., Sebastià, M. T., et al. (2015). Nitrogen yield advantage from grass–legume mixtures is robust over a wide range of legume proportions and environmental conditions. *Glob. Chang. Biol.* 21, 2424–2438. doi: 10.1111/gcb.12880
- Tardieu, F., Reymond, M., Muller, B., Granier, C., Simonneau, T., Sadok, W., et al. (2005). Linking physiological and genetic analyses of the control of leaf growth under changing environmental conditions. *Crop Past. Sci.* 56, 937–946. doi: 10.1071/ar05156
- Taylor, N. L., and Quesenberry, K. H. (1996). *Red Clover Science*, Vol. 28. Dordrecht: Springer Science and Business Media.
- Thomas, R. G. (1987). “The structure of the mature plant,” in *White Clover*, eds M. J. Baker and W. M. Williams (Wallingford: CAB International), 1–29.
- Thomas, R. G. (2003). “Comparative growth forms of dryland forage legumes,” in *Legumes for Dryland Pastures, Proceedings of a New Zealand Grassland Association Symposium*, ed D. J. Moot (Palmerston North: New Zealand Grassland Association), 18–19.
- Thomas, R. G., Li, F. Y., and Hay, M. J. M. (2014). Differential bud activation by a net positive root signal explains branching phenotype in prostrate clonal herbs: a model. *J. Exp. Bot.* 65, 673–682. doi: 10.1093/jxb/ert427
- Turc, O., and Lecoœur, J. (1997). Leaf primordium initiation and expanded leaf production are co-ordinated through similar response to air temperature in Pea (*Pisum sativum* L.). *Ann. Bot.* 80, 265–273. doi: 10.1006/anbo.1996.0431
- Valladares, F., and Niinemets, U. (2007). “The architecture of plant crowns: from design rules to light capture and performance,” in *Handbook of Functional Plant Ecology*, eds F. I. Pugnaire and F. Valladares (Boca Raton, FL: CRC Press), 101–149.
- Van Minnebruggen, A., Cnops, G., Saracutu, O., Goormachtig, S., Van Bockstaele, E., Roldán-Ruiz, I., et al. (2014). Processes underlying branching differences in fodder crops. *Euphytica* 195, 301–313. doi: 10.1007/s10681-013-0997-9
- Van Minnebruggen, A., Roldán-Ruiz, I., Van Bockstaele, E., Haesaert, G., and Cnops, G. (2015). The relationship between architectural characteristics and regrowth in *Trifolium pratense* (red clover). *Grass Forage Sci.* 70, 507–518. doi: 10.1111/gfs.12138
- Vertès, F., Jeuffroy, M. H., Louarn, G., Voisin, A. S., and Justes, E. (2015). Légumineuses et prairies temporaires: des fournitures d’azote pour les rotations. *Fourrages* 223, 221–232. Available online at: <http://www.afpf-asso.fr/index/action/page/id/33/title/Les-articles/article/2054>
- Villani, P. J., and Demason, D. A. (2000). Roles of the Af and Tl genes in pea leaf morphogenesis: shoot ontogeny and leaf development in the heterozygotes. *Ann. Bot.* 85, 123–135. doi: 10.1006/anbo.1999.1004
- White, J. (1979). The plant as a metapopulation. *Annu. Rev. Ecol. Syst.* 10, 109–145. doi: 10.1146/annurev.es.10.110179.000545
- Zaka, S., Ahmed, L. Q., Escobar-Gutiérrez, A. J., Gastal, F., Julier, B., and Louarn, G. (2017). How variable are non-linear developmental responses to temperature in two perennial forage species? *Agric. Forest Meteorol.* 232, 433–442. doi: 10.1016/j.agrformet.2016.10.004
- Zhu, J., Andrieu, B., Vos, J., van der Werf, W., Fournier, C., and Evers, J. B. (2014). Towards modelling the flexible timing of shoot development: simulation of maize organogenesis based on coordination within and between phytomers. *Ann. Bot.* 114, 753–762. doi: 10.1093/aob/mcu051

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Faverjon, Escobar-Gutiérrez, Litrico and Louarn. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Combined Comparative Transcriptomic, Metabolomic, and Anatomical Analyses of Two Key Domestication Traits: Pod Dehiscence and Seed Dormancy in Pea (*Pisum* sp.)

Iveta Hradilová¹, Oldřich Trněný^{2,3}, Markéta Váľková^{4,5}, Monika Cechová^{4,5}, Anna Janská⁶, Lenka Prokešová⁷, Khan Aamir⁸, Nicolas Krezdorn⁹, Björn Rotter⁹, Peter Winter⁹, Rajeev K. Varshney⁸, Aleš Soukup⁶, Petr Bednář^{4,5}, Pavel Hanáček² and Petr Smýkal^{1*}

¹ Department of Botany, Palacký University in Olomouc, Olomouc, Czechia, ² Department of Plant Biology, Mendel University in Brno, Brno, Czechia, ³ Agricultural Research, Ltd., Troubsko, Czechia, ⁴ Department of Analytical Chemistry, Regional Centre of Advanced Technologies and Materials, Palacký University in Olomouc, Olomouc, Czechia, ⁵ Faculty of Science, Palacký University in Olomouc, Olomouc, Czechia, ⁶ Department of Experimental Plant Biology, Charles University, Prague, Czechia, ⁷ Department of Crop Science, Breeding and Plant Medicine, Mendel University in Brno, Brno, Czechia, ⁸ Research Program-Genetic Gains, ICRISAT, Hyderabad, India, ⁹ GenXPro, Frankfurt, Germany

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Ambuj Bhushan Jha,
University of Saskatchewan, Canada
R. Varma Penmetsa,
University of California, Davis, USA

*Correspondence:

Petr Smýkal
petr.smykal@upol.cz

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 23 December 2016

Accepted: 27 March 2017

Published: 25 April 2017

Citation:

Hradilová I, Trněný O, Váľková M, Cechová M, Janská A, Prokešová L, Aamir K, Krezdorn N, Rotter B, Winter P, Varshney RK, Soukup A, Bednář P, Hanáček P and Smýkal P (2017) A Combined Comparative Transcriptomic, Metabolomic, and Anatomical Analyses of Two Key Domestication Traits: Pod Dehiscence and Seed Dormancy in Pea (*Pisum* sp.). *Front. Plant Sci.* 8:542. doi: 10.3389/fpls.2017.00542

The origin of the agriculture was one of the turning points in human history, and a central part of this was the evolution of new plant forms, domesticated crops. Seed dispersal and germination are two key traits which have been selected to facilitate cultivation and harvesting of crops. The objective of this study was to analyze anatomical structure of seed coat and pod, identify metabolic compounds associated with water-impermeable seed coat and differentially expressed genes involved in pea seed dormancy and pod dehiscence. Comparative anatomical, metabolomics, and transcriptomic analyses were carried out on wild dormant, dehiscent *Pisum elatius* (JI64, VIR320) and cultivated, indehiscent *Pisum sativum* non-dormant (JI92, Cameor) and recombinant inbred lines (RILs). Considerable differences were found in texture of testa surface, length of macrosclereids, and seed coat thickness. Histochemical and biochemical analyses indicated genotype related variation in composition and heterogeneity of seed coat cell walls within macrosclereids. Liquid chromatography–electrospray ionization/mass spectrometry and Laser desorption/ionization–mass spectrometry of separated seed coats revealed significantly higher contents of proanthocyanidins (dimer and trimer of galocatechin), quercetin, and myricetin rhamnosides and hydroxylated fatty acids in dormant compared to non-dormant genotypes. Bulk Segregant Analysis coupled to high throughput RNA sequencing resulted in identification of 770 and 148 differentially expressed genes between dormant and non-dormant seeds or dehiscent and indehiscent pods, respectively. The expression of 14 selected dormancy-related genes was studied by qRT-PCR. Of these, expression pattern of four genes: porin (MACE-S082), peroxisomal membrane PEX14-like protein (MACE-S108), 4-coumarate CoA ligase (MACE-S131), and UDP-glucosyl transferase (MACE-S139) was in

agreement in all four genotypes with Massive analysis of cDNA Ends (MACE) data. In case of pod dehiscence, the analysis of two candidate genes (*SHATTERING* and *SHATTERPROOF*) and three out of 20 MACE identified genes (MACE-P004, MACE-P013, MACE-P015) showed down-expression in dorsal and ventral pod suture of indehiscent genotypes. Moreover, MACE-P015, the homolog of peptidoglycan-binding domain or proline-rich extensin-like protein mapped correctly to predicted *Dpo1* locus on PsLGIII. This integrated analysis of the seed coat in wild and cultivated pea provides new insight as well as raises new questions associated with domestication and seed dormancy and pod dehiscence.

Keywords: domestication, legumes, pea (*Pisum sativum*), metabolites, pod dehiscence, seed dormancy, seed coat, transcriptomics

INTRODUCTION

The origin of the agriculture was one of key points in human history, and a central part of this was the evolution of new plant forms, domesticated crops (Meyer et al., 2012; Fuller et al., 2014). The transformation of wild plants into crop plants can be viewed as an accelerated evolution, representing adaptations to cultivation and human harvesting, accompanied by genetic changes (Lenser and Theißen, 2013; Olsen and Wendel, 2013; Shi and Lai, 2015). Common set of traits have been recorded for unrelated crops (Hammer, 1984; Zohary and Hopf, 2000; Lenser and Theißen, 2013). These include loss of germination inhibition and loss of natural seed dispersal (Fuller and Allaby, 2009). The identity of some responsible genes has been revealed (reviewed in Meyer and Purugganan, 2013) through association mapping and genome sequencing, for example in soybean (Zhou et al., 2015), chickpea (Bajaj et al., 2015; Kujur et al., 2015), and common bean (Schmutz et al., 2014).

Members of the *Fabaceae* family have been domesticated in parallel with cereals (Smartt, 1990; Zohary and Hopf, 2000) or possibly even earlier (Kislev and Bar-Yosef, 1988) resulting in largest number of domesticates per plant family (Smýkal et al., 2015). Despite of crucial position of legumes, as protein crops, in human diet as well as crop rotation systems (Foyer et al., 2016), comparably little is known on their domestication. Pea (*Pisum sativum* L.) is one of the world's oldest domesticated crops and is still globally important grain legume crop (Smýkal et al., 2012, 2015). Experimental cultivation of wild peas have demonstrated that both seed dormancy and pod dehiscence cause poor crop establishment via reduced germination as well as dramatic yield losses via seed shattering (Abbo et al., 2011). The loss of fruit shattering has been under selection in the most seed crops, to facilitate seed harvest (Fuller and Allaby, 2009; Purugganan and Fuller, 2009), while in wild plants, shattering is a fundamental trait to assure seed dispersal (Bennett et al., 2011). Orthologous genes and functions were found to be conserved for seed shattering mechanisms between mono and dicotyledonous plants (Konishi et al., 2006). Recently, two genes have been identified to be involved in pod dehiscence in soybean. One of them is the dirigent-like protein (*Pdh1*) promoting pod dehiscence by increasing the torsion of dried pod walls, which serves as a driving force for pod dehiscence under low humidity

(Funatsuki et al., 2014). The functional gene *Pdh1* was highly expressed in the lignin-rich inner sclerenchyma of pod walls. Yet, another NAC family gene *SHATTERING1-5* (Dong et al., 2014) activates secondary wall biosynthesis and promotes the significant thickening of fiber cap cells of the pod ventral suture secondary walls. The differences between wild and cultivated soybean is within promoter region and subsequently expression level (Dong et al., 2014).

Timing of seed germination is one of the key steps in plant life. Seed dormancy is considered as a block to the completion of germination of an intact viable seed under favorable conditions (Baskin and Baskin, 2004; Weitbrecht et al., 2011). In the wild, many seeds will only germinate after certain conditions have passed, or after the seed coat is physically disrupted (Bewley, 1997; Baskin et al., 2000; Finch-Savage and Leubner-Metzger, 2006; Bewley et al., 2013). In contrast, crops were selected to germinate as soon as they are wet and planted (Weitbrecht et al., 2011). Moreover, easy seed imbibition has crucial role in cooking ability of most grain legumes. Hence, reducing seed coat thickness led to a concurrent reduction of seed coat impermeability during the domestication (Smýkal et al., 2014). Seed dormancy had played a significant role in evolution and adaptation of plants, as it determines the outset of a new generation (Nonogaki, 2014; Smýkal et al., 2014). A diverse dormancy mechanisms has evolved in keeping with the diversity of climates and habitats (Nikolaeva, 1969; Baskin and Baskin, 2004; Finch-Savage and Leubner-Metzger, 2006). In contrast to hormone mediated seed dormancy extensively studied in *Arabidopsis* or cereals, we have very little knowledge on physical dormancy, as found in legumes (Baskin and Baskin, 2004; Graeber et al., 2012; Radchuk and Borisjuk, 2014). Although hard-seededness was largely overcome in all domesticated grain legumes except of fodder legumes (Werker et al., 1979; Smartt, 1990; Weeden, 2007), it appears in lentil or soybean depending on the cultivation conditions. Physical seed dormancy is caused by one or more water-impermeable cell layers in seed coat (Baskin et al., 2000; Koizumi et al., 2008; Weitbrecht et al., 2011; Radchuk and Borisjuk, 2014; Smýkal et al., 2014). Numerous *transparent testa* (*tt*) and *tannin deficient seed* (*tds*) mutants (Appelhaagen et al., 2014) indicates the important role of proanthocyanidins and flavonoid pigments in *Arabidopsis* (Graeber et al., 2012) and *Medicago* (Liu et al., 2014) testa development. In *Arabidopsis*

and *Melilotus*, seed permeability is altered due to in mutation in extracellular lipid biosynthesis (Beisson et al., 2007). Similarly, in the *M. truncatula* transcriptomic data set (Verdier et al., 2013a), four of 12 Glycerol-3-phosphate acyltransferases (GPAT) genes were identified as putative orthologs of those reported in soybean (Ranathunge et al., 2010). Furthermore, cells of the outer integument in *M. truncatula* and pea showed abundant accumulation of polyphenolic compounds; which upon oxidation may impact seed permeability (Marbach and Mayer, 1974; Werker et al., 1979; Moïse et al., 2005). Seed dormancy was identified as monogenic trait in mungbean (Isemura et al., 2012); while six QTLs were detected in yardlong and rice bean (Kongjaimun et al., 2012). In pea, Weeden (2007) has identified two to three loci involved in seed dormancy, via testa thickness and structure of testa surface. Two genes involved in seed coat water permeability were recently identified in soybean. One of them, *GmHs1-1*, encodes a calcineurin-like metallophosphoesterase transmembrane protein, which is primarily expressed in the Malpighian layer (macroscleireids) of the seed coat and is associated with calcium content (Sun et al., 2015). Independently of this, *qHS1*, a quantitative trait locus for hardseededness in soybean, was identified as endo-1,4- β -glucanase (Jang et al., 2015). This genes seems to be involved in the accumulation of β -1,4-glucan derivatives that reinforce the impermeability of seed coats in soybean. Interestingly, both genes are positioned closely to each other of soybean chromosome 2.

Development of pea and particularly model legume *Medicago truncatula* seeds have been well-characterized at anatomical (Hedley et al., 1986; Wang and Grusak, 2005) and also transcriptomic and proteomic levels (Gallardo et al., 2007; Verdier et al., 2013a). RNA sequencing (RNA-seq) was used to study changes in gene expression, including *M. truncatula* (Benedito et al., 2008), *Medicago sativa* (Zhang et al., 2015), soybean (Severin et al., 2010; Patil et al., 2015), faba bean (Kaur et al., 2012), *Lotus japonicus* (Verdier et al., 2013b) and chickpea (Pradhan et al., 2014). In pea, transcriptome studies involved vegetative tissues (Franssen et al., 2011), including pods and seeds (Kaur et al., 2012; Duarte et al., 2014; Liu et al., 2015; Sudheesh et al., 2015), and nodules (Zhukov et al., 2015). Seed coat transcriptome of pea cultivars was analyzed in relation to proanthocyanidin pathway (Ferraro et al., 2014) and seed aging (Chen et al., 2013). Moreover, there is pea RNA-seq gene atlas for 20 cDNA libraries including different developmental stages and nutritive conditions (Alves-Carvalho et al., 2015). Comparative transcriptomics study in relation to domestication trait was conducted recently by Zou et al. (2015) in relation to glume and threshing in wheat. Some of the down-regulated genes in domesticated wheat were related to the biosynthetic pathways that apparently define the mechanical strength of the glumes, such as cell wall, lignin, pectin, and wax biosynthesis. Several of so far identified genes underlying key domestication traits (reviewed in Meyer and Purugganan, 2013) are regulated at transcriptional level with altered spatial and temporal expression, such as seed-shattering (*qSH1*) locus disrupting the development of the abscission zone between grains and pedicles in rice (Konishi et al., 2006) or *teosinte branched* (*tb1*) gene causing single stem growth in maize crop (Doebley et al., 1997).

In the present study, we used comparative transcriptomic, anatomical, and metabolite analysis to detect the differences in gene expression, seed coat structure, and metabolites composition between wild and domesticated pea seed coats in relation to one of the two key domestication traits: seed coat and transcriptomic and anatomical analyses of pod dehiscence.

MATERIALS AND METHODS

Plant Material

Four parental genotypes included wild *P. elatius* JI64 from Turkey and cultivated Afghan landrace *P. sativum* JI92 both from John Innes Pisum Collection (Norwich, UK); wild *P. elatius* VIR320 (Bogdanova et al., 2012) from Vavilov Institute Research of Plant Industry (St. Petersburg, Russia) and cultivated *P. sativum* cv. Cameor from INRA France. Furthermore, 126 F_{5,6} recombinant inbred lines (RILs) derived from JI64 and JI92 cross (North et al., 1989) were used to establish respective phenotypically contrasting (dormant vs. non-dormant, dehiscent vs. indehiscent) bulks. *P. elatius* VIR320 differs from other wild peas in relation to the absence of gritty and testa pigmentation, possibly as the result of being either semi-domesticated or hybrid between wild and cultivated pea with unknown origin (Bogdanova et al., 2012).

Seed Water Uptake and Germination Assays

The seeds of four parental and 126 RILs of F₆ generation of JI64 (wild) \times JI92 (cultivated) and reciprocal (RILs) were harvested from glasshouse grown plants (February–May 2015). Twenty five seeds per line were incubated in petri dishes (9 cm diameter) over two layers of medium speed qualitative filter papers (Whatman, grade 1) wetted with 3 ml of tap water and incubated in a 25°C incubator with darkness. Imbibition was scored at 24 h intervals based on changes in seed swelling and germination was determined based on the radicle breaking through seed coat. The percentage, Mean germination time (MGT), Timson index (TI), and Coefficient of Velocity (CV) were calculated over 7 days period. We have used these various mathematical measurements in order to more precisely describe germination process as shown by Ranal and Santana (2006).

Determination of Pod Dehiscence

Pod dehiscence was measured either by direct observation of the pods on the plant or by drying harvested pods at room temperature (Weeden et al., 2002; Weeden, 2007). In case of parental lines JI92 (domesticated, indehiscent pod) and JI64 (wild, dehiscent pod) are dehiscence/indehiscence obvious after pod maturing. On the other hand evaluation of RILs was difficult in some cases, as slight pressure on pods by fingers is necessary for opening. If slight pressure was enough to complete fruit opening the line was evaluated as dehiscent, if not this line was evaluated as indehiscent.

Anatomical Analyses

Samples of seed coat (JI64, VIR320, Cameor and JI92; at least five seeds per genotype) were dissected from dry seed and

saturated with 2% sucrose solution under vacuum. Equal volume of cryo-gel (Cryomatrix Shandon) was added to samples and shake overnight. Saturated samples were mounted into cryo-gel on the alum chuck, frozen down to -25°C and cut in cryotome (Shandon SME, Astmoor, UK) into 12 μm transversal section (Soukup and Tylová, 2014). Sections were stained with toluidine blue (0.01%, w/v in water), alcian blue (0.1%, w/v in 3% acetic acid), aniline blue fluorochrome (Sirofluor; 0.01%, w/v in 100 mM K_2HPO_4 with pH 9), or Sudan Red 7B (0.01%, w/v) according to Soukup (2014). The presence of proanthocyanidins was evaluated by staining with vanillin (Gardner, 1975) and DMACA (Li et al., 1996). Callose immunodetection was performed according to Soukup (2014) using primary antibody toward (1,3)- β -glucan (1:100; Biosupplies Australia PTY Ltd) and anti-mouse IgG Alexa Fluor 488 secondary antibody (1:500; Invitrogen). Control samples were processed without the primary antibody. Sections were observed with an Olympus BX 51 microscope (Olympus Corp., Tokyo, Japan) in bright field, blue (Olympus WB filter—callose immunodetection) or UV (Olympus WU filter—aniline blue fluorochrome and DMACA) excitation. Unstained control sections were surveyed in bright field or UV-excited autofluorescence. Figures were documented with an Apogee U4000 digital camera (Apogee Imaging Systems, Inc., Roseville, CA, USA). Dry intact seeds were vacuum dried and gold coated (Sputter Coater SCD 050, Bal-Tec) before imaging with scanning electron microscope (JSM-6380LV; JEOL, Tokyo, Japan). Twenty days (after flowering) old pods of parental lines (JI64 and JI92) as well as RILs of F6 were fixed in 2% formaldehyde and stored in 4°C for latter observation of pod suture. Samples were cut on hand microtomes at thickness of 100 μm , and the resulting segments were stained 1% phloroglucinol (Sigma, USA) in 12% HCl (Soukup, 2014).

Liquid Chromatography–Electrospray Ionization/Mass Spectrometry (LC/ESI-MS) Analysis

Testa was separated from the rest of the seed, crushed, and extracted using mixture of acetone:water (70:30, v/v) with addition of 0.1% ascorbic acid to achieve efficient extraction of polyphenolic compounds in wide range of polarity and structural diversity (adapted from Amarowicz et al., 2009). 0.5 ml of extract was dried under a stream of nitrogen and solid residue was dissolved in 0.5 ml of methanol. The samples were then analyzed by ultra-performance liquid chromatograph Acquity UPLC I-Class coupled to high resolution tandem mass spectrometer Synapt G2-S with ion mobility separation capability (Waters, Milford, USA). Chromatographic column Raptor ARC-18 (100 \times 2.1 mm, dp = 2.7 μm , Restek) and mobile phases (MP) A: water + 0.1% formic acid, B: acetonitrile + 0.1% formic acid was used for separation of components present in seed coat extracts. Flow rate of mobile phase 0.2 ml/min was applied. Electrospray was used as ion source. Spray voltage 2.5 kV in positive and 1.5 kV in negative ion mode, were used, respectively. Process of the LC/ESI-MS method optimization and detailed setup of mass spectrometer will be provided in Válková et al. (in preparation).

Laser Desorption/Ionization–Mass Spectrometry (LDI-MS) Analysis

Seeds of each genotype/line were mechanically disrupted and the seed coats were separated and pooled (four seeds per genotype). Description of the studied RILs is given in **Table S1**. Small pieces ~ 2 mm were fixed on MALDI plate using a common double sided adhesive tape. The samples were analyzed directly without application of a matrix. The prepared samples were analyzed using high resolution tandem mass spectrometer Synapt G2-S (Waters) equipped with vacuum MALDI ion source. For desorption/ionization a 350 nm 1 kHz Nd:YAG solid state laser was used. Details of LDI-MS setup and analytical parameters of hydroxylated fatty acids can be found in Cechová et al. (in preparation).

Metabolite Data Treatment

The obtained LC/ESI-MS and LDI-MS data were processed by MarkerLynx XS a software extension of MassLynx platform (Waters). The processed data matrix, i.e., after extraction, normalization and alignment of retention times (in case of LC/ESI-MS data), m/z-values and intensities of signals, were transferred to Extended Statistics (XS) module, EZinfo (Umetrics, Malmo, Sweden), and studied by principal component analysis (PCA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA). Both PCA and OPLS-DA were used for reduction of data dimensionality. OPLS-DA is a multivariate statistical method employing latent variable regression developed as an extension of more frequently used partial least squares method (Trygg and Wold, 2002). Coordinates of particular samples and RT_m/z pairs (or m/z-values in the case of LDI-MS data) in appropriate biplots and S-plots were used for evaluation of dormant and non-dormant genotypes mutual segregation and significance of detected signals of metabolites. The procedure was adopted and modified from Kučera et al. (2017). The most significant markers were further studied by targeted MS/MS experiments to reveal their identity (Cechová et al., in preparation; Válková et al., in preparation).

RNA Isolation

For Massive analysis of cDNA Ends (MACE) each sample of parental genotype was composed by several pooled developmental stages of seed coat (2, 3, 4 weeks and older) as it is unknown at which stage putative candidate genes are expressed. Seven selected RILs forming each of dormant resp. non-dormant bulk were previously (at F₆ generation) and after mature seed harvest (of F₇ generation used for RNA isolation) tested for germination behavior (**Table S1**). Seed coat were dissected under stereomicroscope, immediately frozen in liquid nitrogen and stored at -70°C until use. Frozen seed coats or dorsal and ventral sutures of pods were ground to a fine powder with liquid nitrogen using sterile mortars and pestles. Total RNA was isolated from seed coat (~ 100 mg) using the BioTeke Plant Total RNA Extraction Kit (China) or NucleoSpin RNA Plant kit (Macherey Nagel) for pods, according to the manufacturer's instructions. Yield/quantity and purity was determined by using NanoDrop 2000 spectrophotometer (Thermo Scientific) and diluted in DEPC- H_2O to 100 ng/ μl . Isolated RNAs were treated

with DNase according to Baseline-ZERO™ DNase protocol (Epicenter). In case of parental genotypes (JI64, JI92, Cameor, and VIR320) four consecutive developmental stages of seeds (14–37 DPA) were taken each represented by 1.25 µg of total RNA. The RIL bulks were made of 1.425 µg of total RNA of each of seven lines. In case of RNA samples used for pod dehiscence study, two parental lines (JI64 and JI92) and two bulks of contrasting RILs (with dehiscent or indehiscent pods) were used. The bulk of dehiscent RILs was established from eight and bulk of indehiscent RILs from five lines using excised pod sutures of 10 and 20 days after flowering. Each of these four final samples contained ~1 µg of total RNA each. The integrity of the RNA samples was examined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, USA).

Massive Analysis of cDNA Ends (MACE)

MACE libraries were generated using GenXPro's MACE kit (GenXPro GmbH, Frankfurt, Germany) as described in Zawada et al. (2014). Briefly, cDNA from 5 µg of total RNA was randomly fragmented and biotinylated 3' ends were captured after binding to a streptavidin matrix. A library ready for high-throughput sequencing was prepared using TrueQuant adapters included in the kit. The library consisted of 50–700 bp-long fragments derived from the 3'-end of the cDNAs. The 5'-ends of the libraries were sequenced on a HiSeq 2000 machine (Illumina) with 100 cycles to generate the MACE tags, each tag representing one single transcript molecule. In total, 6 cDNA libraries were prepared and sequenced for seed dormancy, while 4 libraries for parents and two contrasting bulks for pod dehiscence, each providing over 10 million reads (Table S2).

Bioinformatics

After sequencing the reads are in 5'–3' orientation. To remove PCR-bias, all duplicate reads detected by the TrueQuant technology were removed from the raw datasets. Low quality sequence-bases were removed by the software Cutadapt (<https://github.com/marcelm/cutadapt/>) and poly(A)-tails were clipped by an in-house python-script. The reads were aligned to reference sequences using Novoalign (<http://www.novocraft.com>). This tool maps reads to reference sequences depending on certain parameters (i.e., quality) and calculates thresholds for each assignment. The reference sequences consisted of all *Pisum* mRNA sequences from NCBI. We annotated these sequences to all Fabaceae proteins from Uniprot "<http://www.uniprot.org/>" by BLASTX to Swissprot ("sp|...", good annotation) and afterwards to Trembl ("tr|...", less good annotation) protein sequences. All reads that could not be mapped to *Pisum* mRNA sequences from NCBI were used for a *de novo* assembly to generate contigs denoted as "noHitAssembly_xxx" and annotated in the same way as the *Pisum* mRNA sequences from NCBI. Normalization and test for differential gene expression between the bulks were calculated using the DEGSeq R/Bioconductor package (Wang et al., 2010). Differential gene expression was quantified as the log₂ ratio of the normalized values between two libraries (log₂ FC). The *p*-value and correction for multiple testing with the Benjamini–Hochberg false discovery rate (FDR) were computed due to determining significance of gene expression differences in pairwise comparisons of libraries. Lists of Differentially

Expressed Genes (DEGs) for three comparisons of contrasting phenotype (dormancy: wild × cultivated, RILs and their parents, dehiscence: RILs and their parents) were made based on combination of pairwise comparisons of log₂ FC ratio of the normalized values (log₂ FC > 2, log₂ FC < –2) and FDR (FDR < 0.01) between all libraries of these groups.

GO and KEGG Annotation

Gene Ontology (GO) enrichment analysis and normalized gene expression data were used to identify function and relationships of differentially expressed genes (Young et al., 2010). The results of the GO analysis were then exported into the Blast2go for the final annotations. The annotations provided the fragments with blast hit with the appropriate gene ontology terms which were classified into three categories: biological process (BP), cellular components (CC), and molecular function (MF). The DEGs were subjected for their presence in the different Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. The various enzyme activities and the DEGs involved in the KEGG pathways were revealed for each of the combinations.

Genetic Mapping of Dehiscence Specific SNPs-Methodology

Transcripts containing SNP with at least five reads in both samples that are homozygous distributed in dehiscent vs. indehiscent RILs bulks with only one false allele read in 100 in either bulks were considered dehiscence specific. SNPs were discovered using Joint-SNV-Mix (Roth et al., 2012). The output given by Joint-SNV-Mix was furthermore processed by GenXPro in-house software to filter the SNPs. A minimum coverage of 10 bp was needed to be identified as an SNP. To identify the genomic localization of the SNP the surrounding region of the SNP was assigned per blastn to the genome of *M. truncatula* "JCVI.Medtr.v4.20130313" from <http://jcv.org/medicago>. The "snviewer" a webtool from the "<http://tools.genxpro.net>" was used to visualize the data.

Real-Time Quantitative Reverse Transcription PCR

Gene-specific oligonucleotide primers were designed (Table S3) based on MACE consensus sequences using the FastPCR software (Kalendar et al., 2014). The expression of selected candidate genes was validated by quantitative real time PCR (qRT-PCR). RNA samples (treated with DNase) were reverse-transcribed with Oligo(dT)₁₅ primer (Promega) in a two steps reaction in final volume 40 µl. The qRT-PCR analysis was run on the CFX96™ Real-Time Detection System (Bio-Rad) using the SensiFast SYBR® No-ROX kit (Bioline) or LightCycler® 480 SYBR Green I Master kit (Roche) in case of pod dehiscence study. Primers were designed using FastPCR or Oligo Primer Analysis Software (Molecular Biology Insights, USA) and produced amplicons ranging from 77 to 220 bp (Table S3). Every PCR reaction included 2 µl cDNA (1:10 diluted cDNA), 5 µl 2× SensiFAST SYBR mix or LightCycler® 480 SYBR and 400 nM of each primer in final volume 10 µl. The expression was studied at two developmental stages in four contrasting parental genotypes (JI64, JI92, Cameor, and VIR320) and in case of dehiscence study also contrasting RILs with dehiscent or indehiscent pods. The

conditions for PCR were: 95°C for 2 min; 45 cycles of 95°C for 10 s, 55°C for 30 s, and 72°C for 20 s; followed by a melting curve of 65–94°C (recovered every 0.5°C held for 0.5 s). The specificity of primers was confirmed by melting curve and gel analysis of products. Quantification of transcript level was determined by CFX Manager Software (Bio-Rad). Actin or β -tubulin gene was used as a reference to normalize relative quantification (Ferraro et al., 2014) using the comparative Ct ($2^{-\Delta\Delta C_t}$) method. Changes in transcript were estimated as fold change relative to the expression in the genotype Cameor (younger stage) in case of seed dormancy study and genotype JI92 (younger stage) in case of pod dehiscence study.

RESULTS

Seed Coat Mediated Dormancy of Wild Pea Seeds

Anatomical and Germination Differences between Seed Coat of Wild and Cultivated Peas

Wild pea seeds display high level of dormancy mediated by seed coat permeability. Two contrasting parental pairs of wild (*P. elatius*) JI64 and VIR320 and cultivated (*P. sativum*) cv. Cameor and JI92 peas were selected as they differ in testa pigmentation, thickness, and dormancy levels as well as pod dehiscence trait. Moreover, RILs were generated from cross of JI64 and JI92 to facilitate mapping. While cultivated pea seeds imbibe readily and germinate within 24 h (JI92, cv. Cameor), wild pea seeds remain highly dormant and imbibe and germinate at 8% (JI64) or 30% levels (VIR320) after 7 days (Figure 1). Mean germination time, Timson index, and Coefficient of Velocity are also very different between respective parental genotypes, 3.4 (MGT), 1 (TI), and 0.29 (CV) for JI92 while being 7, 0.008, and 0.14 for JI64, respectively. RILs displayed more variability in each of the respective measures (Figure S1), with wide range of percentage of germination (at 7 days) from 4 to 100%, 0.61 to 4.44 CV, 1 to 7 MGT, and 0 to 0.95 TI. All these parameters indicate complexity of imbibition and germination processes as none single is sufficient to fully describe any given line. As shown in Table S1, non-dormant bulk had on average 56% germination over 7 days period, 1.58 CV, 0.36 TI, and 68.02 MGT respective, contrast to dormant bulk lines having 22%, 1.38 CV, 0.28 TI, and 104.71 MGT, respectively. Similarly testa thickness was 107 vs. 135 μ m, respectively.

Testa thickness was analyzed by micrometric, light microscopy or SEM measurements. Especially JI64 and JI92 differ substantially in palisade cells length, which contributes to overall testa thickness (Figure 2). Dormant genotype JI64 has significantly thicker testa, which might contribute to the water impermeability of seed coat of dormant pea genotypes. There are considerable differences in surface pigmentation and texture of individual lines (Figures 3a,d,g,j). While Cameor is not pigmented visible pigmentation is present in other genotypes with different intensity and localization. The texture of surface is variable among genotypes, particularly in details of macrosclereid tips arrangement defining the surface shape being covered by thin cuticle (Figures 3b,c,e,f,h,i,k,l and Figure S2).

The most obvious is gritty surface of JI64 which was absent in all the other genotypes included in this study. The continuity of surface (cuticle integrity) was interrupted locally by minor fissures in all genotypes. Large fissures developed in seed coat of the non-dormant genotypes Cameor and JI92, mostly in the hilar and strophilar region later during the imbibition (data not presented). The cytological arrangement of seed coat of tested genotypes varies particularly in macrosclereids. The surface is covered with thin cuticle (Figure S2) which covers the outer extremities of palisade macrosclereids. Based on light microscopy we did not see any apparent difference in cuticle properties among genotypes as revealed by autofluorescence or Sudan staining (Figure S2). Interestingly, cuticle is not the only lipidic material localized close to the seed coat surface. Non-cuticular lipidic extracellular material was present also in the very tips of the macrosclereids (Figures S2e–h) containing autofluorescent material (Figures S2a–d). The analysis of seed coat surface was complemented with histochemical analysis of selected compounds of cell wall. Metachromatic staining with toluidine blue of non-dormant genotype Cameor exhibited high level of polyanionic cell wall components, while in contrary non-dormant, but well-pigmented JI92 showed lower abundance of polyanionic compounds similarly to dormant genotypes where metachromasy was mostly limited to macrosclereid tips below the cuticle (see Figures S2a–d). Staining with Alcian blue further supports such conclusion (data not shown). Interestingly, metachromatic staining was enhanced with 3% acetic acid pretreatment (Figures S3e–h). Clear connection between presence of tannins and toluidine blue stainability is well-documented in JI92, where deeper staining is present out of pigmented spots (proanthocyanidin positive; Figure S3f). The presence of proanthocyanidin within cell walls of macrosclereids was not detected in non-dormant genotype Cameor but was obvious in JI92 as well as in the dormant genotypes (Figure 4) using both HCl-Vanilin and DMACA tests. Condensed tannin presence was never recorded within the light line (Figure 4) of macrosclereids of any genotype. Macrosclereids of dormant type genotypes seems to be enriched with proanthocyanidins in the cell walls of the entire macrosclereids up to the light line. Seed coat of JI92 is highly enriched with proanthocyanidins only in the dark pigmented spots. Interestingly, the abundance of condensed tannins negatively correlates with toluidine blue stainability and sirofluor staining of cell walls indicating tannin linkages to other compounds within the cell wall. Detected tannins are not extractable with ethanol, or 1M HCl. Alkaline hydrolysis of cell wall bound tannins by 1M sodium hydroxide resulted in loss of tannins stainability (Figures 4c,f). There was intense aniline blue staining of macrosclereids in all genotypes, particularly in the light line of macrosclereids (Figures S2i–l) and their outer part composed in majority of the secondary cell walls. The strongest signal was observed in dormant genotypes, especially in the light line. However, the signal for callose specific antibody did not correspond with aniline blue fluorophore and in general was found rather weak and discretely localized (Figure S2i—inlay). Phloroglucinol staining indicative of lignin provided no response in non-dormant nor in dormant pea genotypes indicating the absence

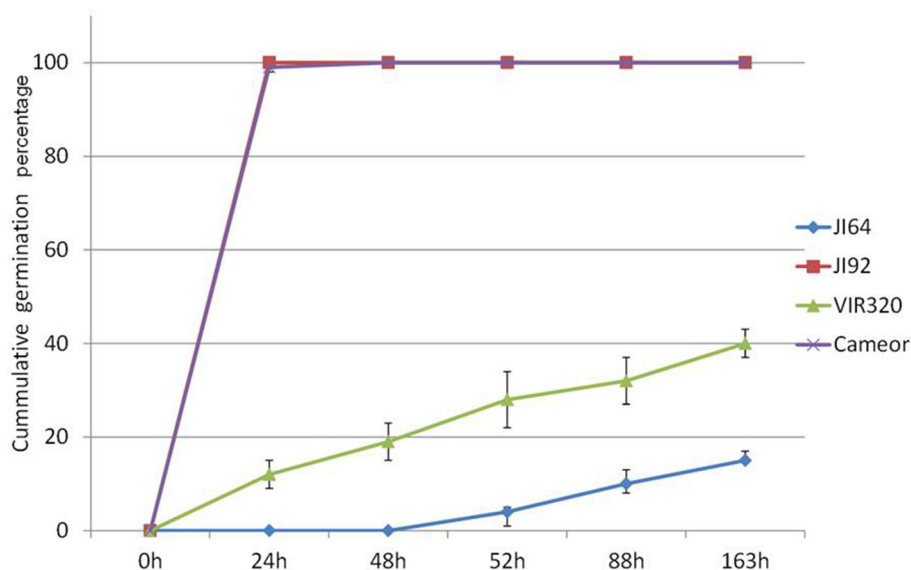


FIGURE 1 | Cumulative germination percentage of wild *P. elatius* (JI64, VIR320) and cultivated *P. sativum* (JI92 and cv. Cameor) seeds tested at 25°C over the period of 163 h.

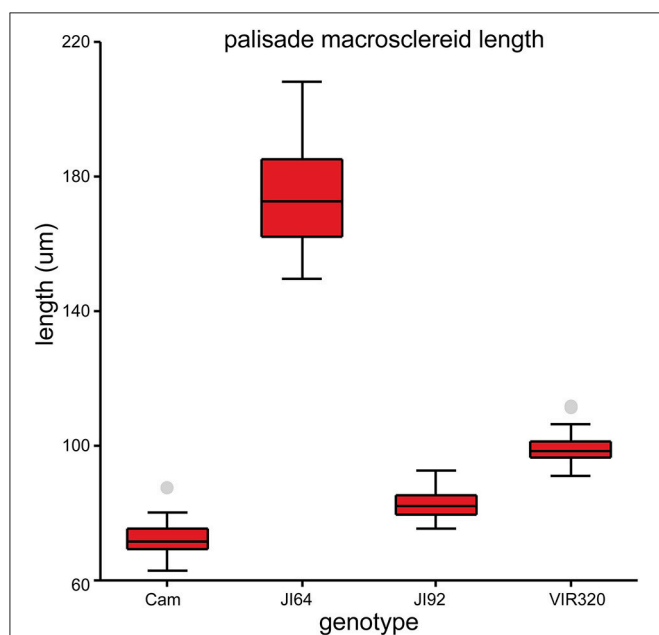


FIGURE 2 | Length of the seed coat palisade cells of selected pea genotypes (Cameor, JI92, JI64, and VIR320). Box plot of median with 25th and 75th percentile, whiskers are 5th and 95th percentile; $n = 75$; each genotype is different from the other (ANOVA $p < 0.001$).

of significant amount of lignins in the testa of analyzed pea genotypes.

Chemical Analysis of Seed Coat Composition

Detection of metabolites present in seed coat related to dormancy was based on comparison of LC/ESI-MS and LDI-MS data of

dormant and non-dormant pea genotypes using principal component analysis and orthogonal projection to latent structures. **Figure 5** reflects the differences in coordinates of particular genotype samples in corresponding Score plot obtained by Principal Component Analysis of LC/ESI-MS data. Although, individual coordinates do not exhibit statistically significant differences among all the genotypes (e.g., $t[2]$ coordinates of Cameor, Terno, and VIR320), location of each genotypes given by combination of both coordinates provided resolution among particular genotypes. The differences in the coordinates (mutual orientations and values) clearly show the separation of dormant (i.e., L100, JI64, and VIR320) from non-dormant genotypes (i.e., Terno, Cameor, and JI 92) using the acetone:water extract. Separation of L100 and JI64 from non-dormant genotypes is much more significant compared to the separation of VIR 320. This can be explained by possible semi-domesticated status of this genotype. Based on the achieved separation of dormant and non-dormant genotypes by unsupervised Principal Component Analysis (PCA), supervised Orthogonal Projection to Latent Structures (OPLS-DA) was used to find signals mostly responsible for the chemical differences in dormant and non-dormant samples. Those signals (m/z -values of markers with increased intensity in dormant genotypes compared to non-dormant ones) were studied in detail by targeted tandem mass spectrometry (study of their fragmentation after collision induced dissociation in collision cell of mass spectrometer). Details of analytical interpretation can be found in Válková et al. (in preparation). Attention was especially focused on the chemical differences between morphologically the most similar pair of genotypes, i.e., JI 64 and JI 92. Combination of information about retention time, exact mass measurement and fragmentation revealed the

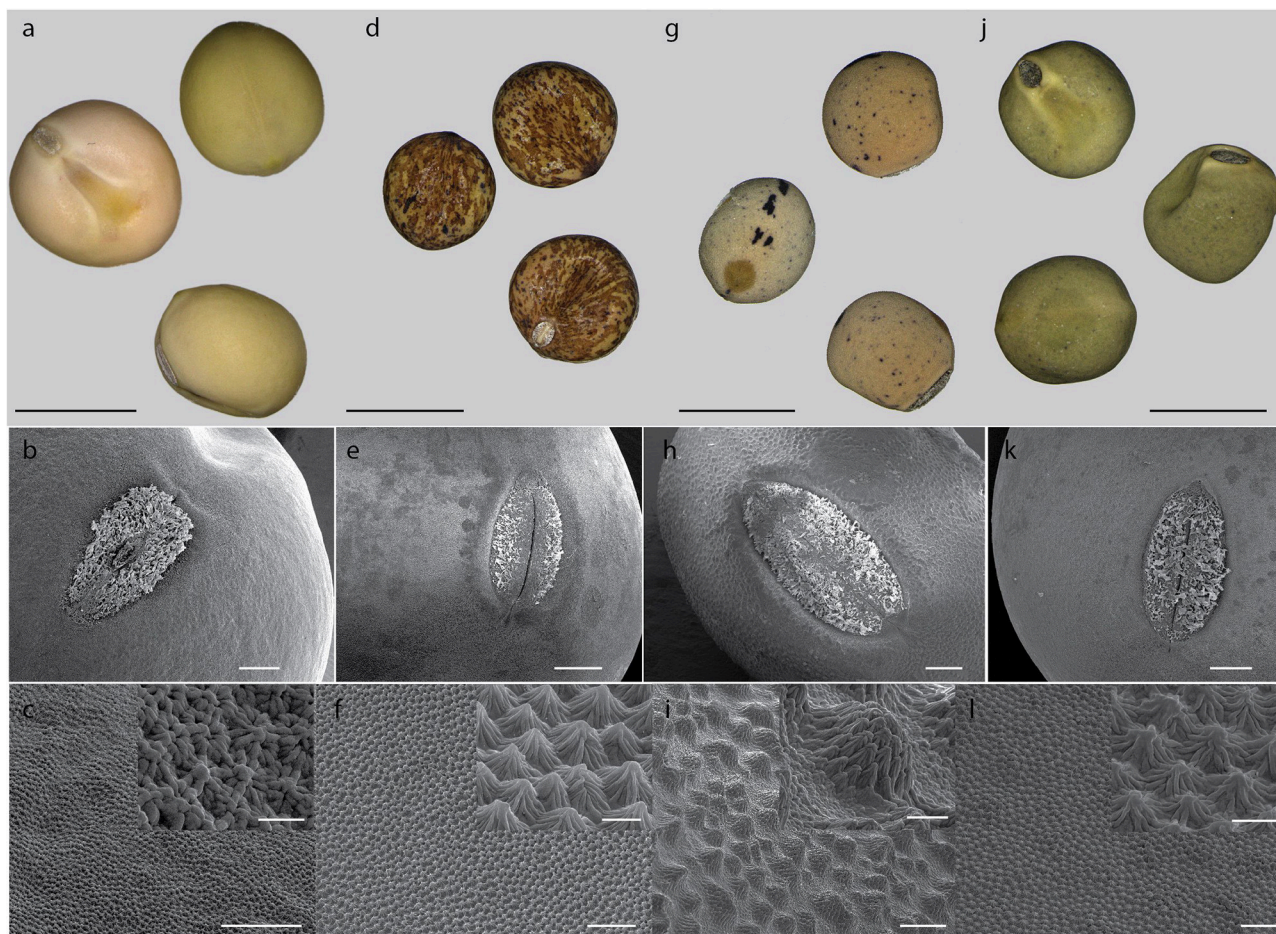


FIGURE 3 | Seed coat surface of selected pea genotypes. Pigmentation of seed coat: Cameor (a), JI92 (d), JI64 (g), and VIR320 (j); scale bar = 5 mm. Surface texture (SEM): Cameor (b,c), JI92 (e,f), JI64 (h,i), and VIR320 (k,l) overall view, scale bar = 500 μ m; details of extrahilar region (c,f,i,l) with inlay of high magnification, scale bars = 100 μ m (c,i) and 50 μ m (f,l), inlay scale bars = 10 μ m (c) 20 μ m (l).

identity of the most significant dormancy markers found in acetone-water extracts—dimer and trimer of gallicocatechin (m/z 611.1387 and 915.1945, deviation of measured from theoretical m/z -value of parent ion, dtm, -0.8 and -3.3 mDa), respectively, quercetin-3-rhamnoside (m/z 449.1045, dtm -3.3 mDa) and myricetin-3-rhamnoside (m/z 465.1112, dtm 7.9 mDa). Analogously, the chemical differences between dormant JI64 and JI92 genotypes were studied by laser desorption-ionization mass spectrometry (LDI-MS). Measurement in negative ion mode in combination with PCA a OPLS-DA revealed marked differences in the profile of particular hydroxylated long chain fatty acids [i.e., m/z 411.3865, hydroxyhexacosanoate (dtm 2.7 mDa); m/z 425.3927, hydroxyheptacosanoate (dtm -6.8 mDa); m/z 427.3875, dihydroxyhexacosanoate (8.8 mDa); m/z 437.4033, hydroxyoctacosanoate (dtm 3.8 mDa); m/z 441.3973, dihydroxyheptacosanoate (2.9 mDa) and m/z 455.4180, dihydroxyoctacosanoate (dtm 8.0 mDa)]. Two orders of magnitude higher normalized signals of dihydroxyheptacosanoate and dihydroxyoctacosanoate were measured in JI 64 compared to JI 92, i.e., $(1.21 \pm 0.92) \cdot 10^{-2}$ and

$(2.50 \pm 1.54) \cdot 10^{-2}$ vs. $(1.34 \pm 0.02) \cdot 10^{-4}$ and $(2.19 \pm 0.58) \cdot 10^{-4}$, respectively. **Figure 6** shows the normalized signals of both dihydroxylated long chain fatty acids in seed coats of JI64, JI92, and RILs with respect to dormancy. The majority of dormant RILs exhibit higher content of those fatty acids compared to non-dormant ones.

Seed Coat Transcriptome Differences between Wild and Cultivated Pea

In order to understand the mechanism of testa mediated dormancy we selected seed coat derived from two contrasting parental pairs and two bulks of RILs differing in dormancy level to extract RNA and to identify candidate genes using the next-generation sequencing method of Massive Amplification of cDNA Ends (MACE). The isolation of RNA from wild pea seed coat tissue proved to be very difficult. It is likely that high due to content of free metabolites (oligosaccharides and proanthocyanidins). We assessed the expression patterns in domesticated (cv. Cameor and JI92 landrace) vs. wild *P. elatius* (JI64, VIR320) pea. Each sample has yielded between 8 and 15

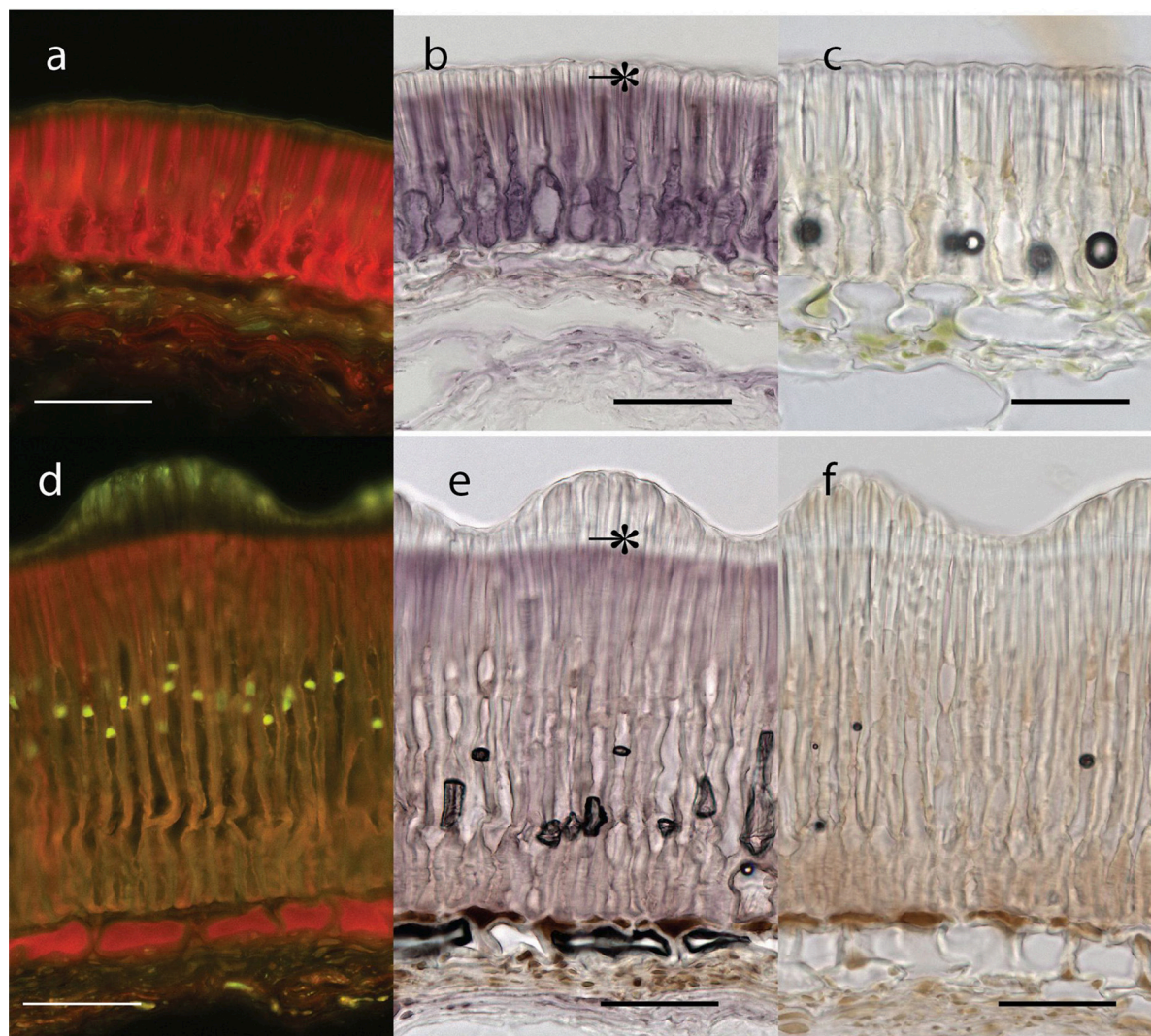
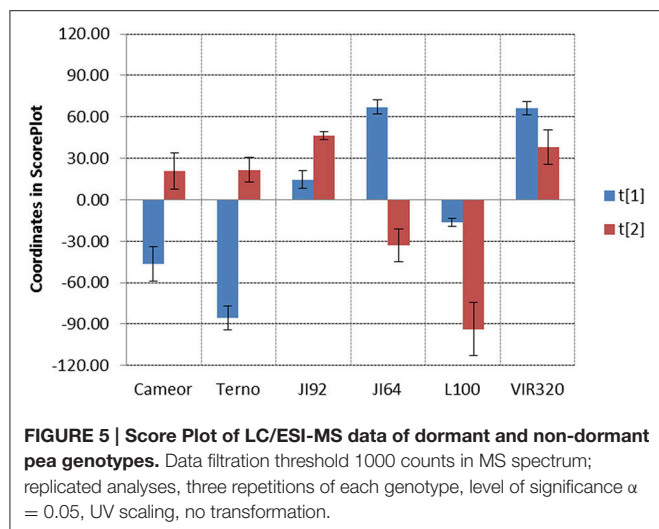


FIGURE 4 | Seed coat transverse sections from extrahilar region stained with DMACA for proanthocyanidins polyphenolics (bar = 50 μm): Upper row—JI92, Lower row—JI64. Blue excited red fluorescence of DMACA (**a,d**) and violet coloration in bright field (**b,e**). Similar sections as in (**c,f**) but stained with DMACA after mild alkaline hydrolysis; scale bars = 50 μm . *, Light line.

million clean reads (**Table S2**). Bioinformatics analysis resulted in identification of 144,000 transcripts (e.g., MACE annotated fragments) expressed in immature seed coat tissue. We have used stringent values of false discovery rate ($\text{FDR} \leq 0.01$) and fold-change ($\log_2 \text{FC} \geq 2$) as a threshold to identify the significant differences in the gene expression. Applying these criteria, a total of 10,132–11,808 transcripts were found differentially expressed between cultivated (JI92, Cameor) and wild (JI64, VIR320) parents (**Table S2**). Of these 770 were differentially expressed between all wild vs. cultivated genotypes, of these 374 were up-regulated in cultivated genotypes, and 396 down-regulated (**Figure 7A**, **Table S4**), when annotated to pea transcriptome, respectively.

A heat-map of 1,000 genes with the highest variance among normalized expression between cultivated and wild pea samples

(**Figure 8A**) further illustrates differences between domesticated and wild pea seed coat expressed genes. This comparison shows the differences between respective pairs (cultivated vs. wild). In addition to contrasting parental genotypes, two bulks of seed coats at several pooled developmental stages of seven dormant and seven non-dormant RILs (**Table S1**) were analyzed. The bulks were included to minimize identification of genes associated with respective genetic background rather than dormancy trait. This is clearly shown when DEGs are compared between parents and RIL bulks (having largest number of specific DEGs between parents e.g., 6,204 up-/5,604 down-regulated transcripts, while only 869 and 1,014 down-regulated transcripts in RIL non-dormant vs. dormant bulks. In case of gene expression profile of dormancy and non-dormancy RILs and their parents, 299 DEGs were found. When comparison



included JI64 and JI92 parents and two respective RIL bulks, there were 83 up- and 216 down-regulated genes (**Figure 7B**). In order to visualize the expression pattern of RILs and their parents, heatmaps were constructed for 11,808 genes from libraries of dormancy and 6,259 genes from libraries of dehiscence (**Figure 8B**).

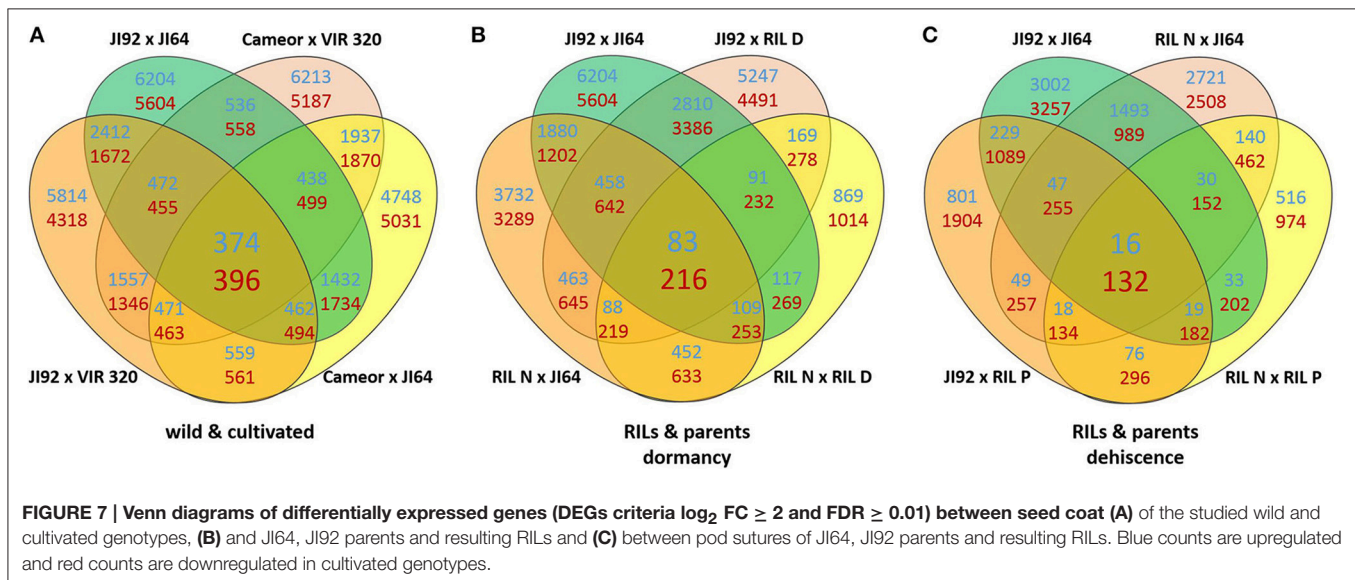
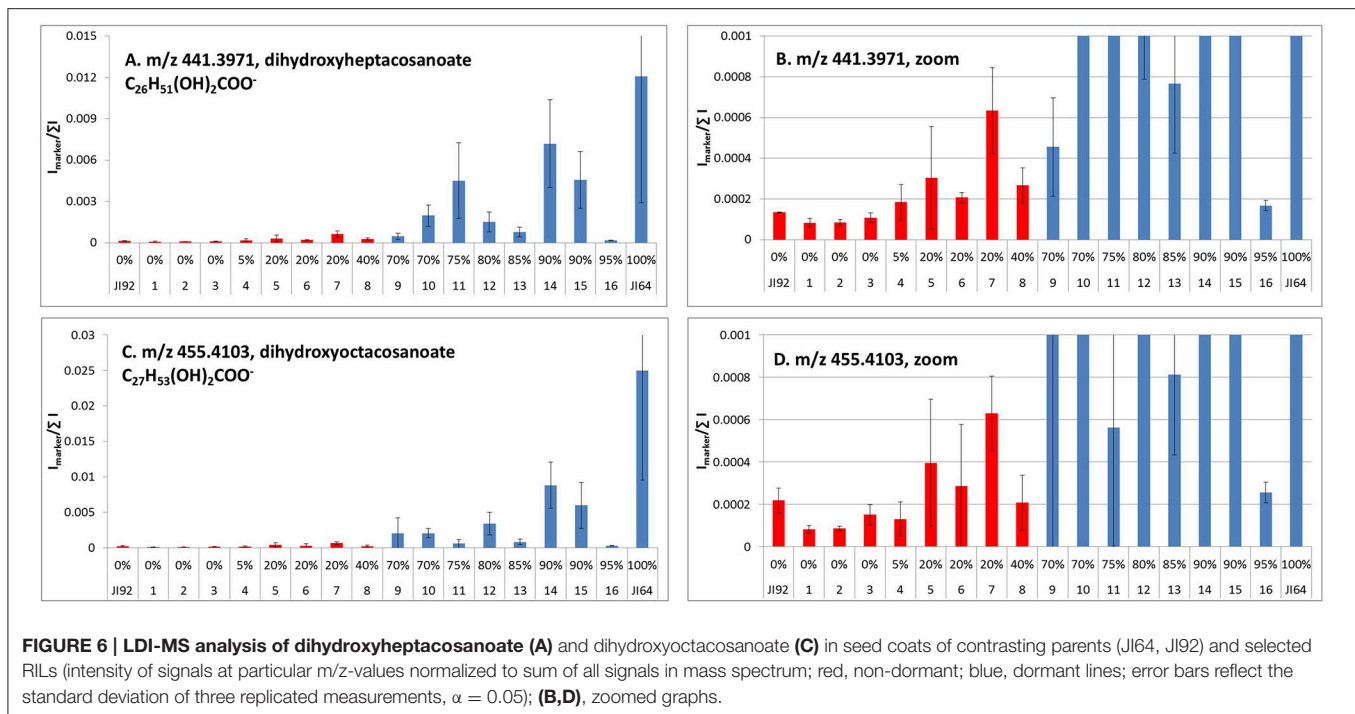
Verification of Differentially Expressed Genes during Seed Coat Development

We selected DEGs based on expression level difference and gene annotation (**Table S4**), regarded as candidate genes for seed coat mediated dormancy in two possible directions of evolutionary changes (i.e., up- or down-regulated in the domesticated pea compared to its wild progenitor). To validate MACE results, expression levels of 14 selected DEGs was analyzed by qRT-PCR. According to MACE results, the selected genes comprised of five up-regulated genes (direction dormant to non-dormant): porin (MACE-S082), NADPH-cytochrome P450 reductase (MACE-S101), peroxisomal membrane PEX14-like protein (MACE-S108), UDP-glucose flavonoid 3-O-glucosyltransferase (MACE-S139), xyloglucan:xyloglucosyl transferase (MACE-S141), and 9 down-regulated genes: NADH-cytochrome b5 reductase (MACE-S019), divergent CRAL/TRIO domain protein (MACE-S066), 1-deoxy-D-xylulose, 5-phosphate, reductoisomerase (MACE-S069), probable aldo-keto reductase 1 (MACE-S070), cupin RmlC-type (MACE-S110), heavy metal transport/detoxification protein (MACE-S111), 4-coumarate CoA ligase (MACE-S131), cytochrome P450 monooxygenase CYP97A10 (MACE-S132), β -amyrin synthase (MACE-S135). Of these tested 14 genes, 4 genes (MACE-S082, MACE-S108, MACE-S131, and MACE-S139) were in agreement in all four genotypes with data obtained with MACE method (**Figure 9**). Relative expression level by MACE-S019 and MACE-S135 was completely contrary to MACE results, where higher values were in wild dormant genotypes. The qRT-PCR expression pattern of MACE-S066 and MACE-S101 were different in two genotypes (Cameor and VIR320) compared to MACE data, with MACE-S066 higher expressed in wild dormant

(VIR320) and MACE-S101 higher expressed in cultivated non-dormant genotype (Cameor). Discrepancy between qRT-PCR and MACE methods in JI64 and JI92 was found in case of MACE-S069, MACE-S070, MACE-S110, MACE-S111, MACE-S132, and MACE-S141.

Enrichment Analysis of DEGs Functional Classes between Wild and Domesticated Pea Seed Coats

In order to investigate transcriptome changes in seed coat associated with evolution under domestication, we assessed the expression patterns of the DEGs in domesticated (cv. Cameor and JI92 landrace) vs. wild *P. elatius* (JI64, VIR320) pea. We identified 770 DEGs (583 respectively, when ambiguous are removed) in seed coat between wild and domesticated peas. Due to the absence of complete pea genome and likely specificity of seed coat tissue, we could annotate 66% of MACE fragments. Moreover, between 36 and 41% produced ambiguous assignment (**Table S2**). For DEGs sequences assigned to GO terms, we observed differences within all three compounds: cellular components, molecular function, and biological process. Several GO groups were found differently expressed. In GO enrichment of DEGs between wild and cultivated the most interesting results belongs to Molecular function group (**Table S5**). The most DEGs were found in phenylpropanoid (17 genes in KEGG pathway) and flavonoid (11 genes) biosynthetic pathways (**Figures S4, S5**). These included O-hydroxycinnamoyltransferase (EC:2.3.1.133), dehydrogenase (EC:1.1.1.195), O-methyltransferase (EC:2.1.1.68), gentiobiase (EC:3.2.1.21), lactoperoxidase (EC:1.11.1.7), ligase (EC:6.2.1.12), reductase (EC:1.2.1.44), and 4-monooxygenase (EC:1.14.13.11) or synthase (EC:2.3.1.74), reductase (EC:1.3.1.77), O-hydroxycinnamoyltransferase (EC:2.3.1.133), isomerase (EC:5.5.1.6), 3'-monooxygenase (EC:1.14.13.21), 4-monooxygenase (EC:1.14.13.11) genes (**Table S5**), respectively. Enzyme 4-coumarate CoA ligase (MACE-S131) catalyzes conversion of 4-coumarate and other derivatives to corresponding esters serving to generate precursors for formation lignin, suberin, flavonoids. In general, main differences in gene expression was detected between enzymes that played important role in secondary metabolites biosynthesis. Different levels of expression were observed for the cellulose synthase enzyme (EC:2.4.1.12) and two enzymes of pectin metabolism pectate lyase (EC:4.2.2.2) and pectin methylesterase (EC:3.1.1.11), which may interfere together with enzymes from the phenylpropanoid and lignin biosynthesis to the structural composition of cell wall. Enzyme 3-hydroxyacyl-CoA (EC:1.1.1.35) dehydrogenase that belongs to fatty acid elongation pathways is also up-regulated in dormant (wild) genotypes. This enzyme participating in of fatty acid biosynthesis showed changes in expression between wild and cultivated genotypes (**Figure 6**). Out of 548 DEGs, 307 (56%) were assigned to GO-term groups, including 171 (56%) down-regulated and 136 (44%) up-regulated in the domesticated pea seed coat compared to wild samples. As shown in **Table S5**, the known DEGs were mainly classified into 40 functional categories and involved in 19 biological processes. The results showed that these DEGs mainly distributed in plasma membranes and nucleus after genes expression, and participated in the biological process



of biosynthetic process (60 genes, 12%), metabolism (183 genes, 36%), regulation of transcription (3 genes, 0.5%), transporting (17 genes, 3%), stress response (16 genes, 3%), cell division and differentiation (180 genes, 35%), localization (30 genes, 6%), establishment of localization (28 genes, 6%), lignin synthesis and so on. Through comparative analysis, the two most abundant sub-classes were biosynthesis processes and metabolic processes. The KEGG pathway analysis showed the presence of many genes PsCam051542, PsCam037704, PsCam043296, PsCam000856, PsCam038256, PsCam049689, PsCam016941, PsCam049689, PsCam050533 involved phenylpropanoid biosynthesis. Similarly, PsCam049689, PsCam038227, PsCam005153, and PsCam050665

were found to be associated with flavonoid biosynthesis (Table S5).

Pea Pod Dehiscence

The structure of pea pericarp follows the common arrangement in *Fabaceae*. The exocarp consists of thick walled epidermis, the relatively thick mesocarp is arranged in several layers of parenchyma and endocarp composed of lignified sclerenchyma on inside of which is thin-walled epidermis. Both exocarp and mesocarp are rich in pectins, as indicated with metachromatic staining of toluidine blue (Figure 11).

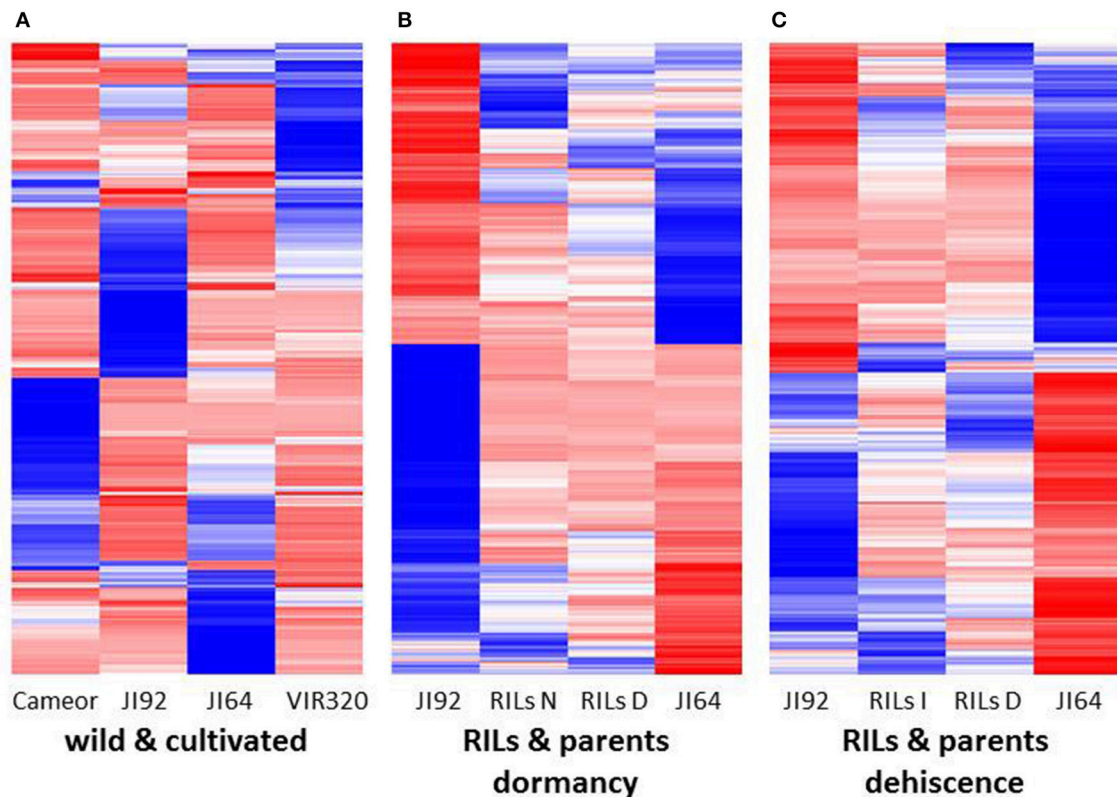


FIGURE 8 | The heat maps **(A)** of 1,000 genes with the highest variance among normalized expression values for cultivated and wild pea samples, **(B)** of 11,808 DEGs for dormancy, and **(C)** of 6,259 DEGs for dehiscence between RILs parents samples in MACE experiments.

Differentially Expressed Genes between Dehiscent and Indehiscent Pods

In the case of pea pod dehiscence, we used MACE methodology to find differences in expression profiles between JI92 (domesticated, indehiscent pod) and JI64 (wild, dehiscent pod) and between two bulks of contrasting RILs. For each sample we used bulk of three developmental stages (2, 4 weeks and older) of dissected pod suture tissue (Table S2). Across all dehiscent and indehiscent libraries 148 DEGs were found (Figure 7C). Of these, 132 DEGs were down-regulated and 16 were up-regulated in indehiscent libraries. For gene expression analysis via qRT-PCR we selected 20 gene candidates with the most different expression in MACE analysis between contrasting lines (dehiscent and indehiscent pod). Nineteen of them were recognized by MACE as down expressed (with lower expression in domesticated indehiscent genotype) and one expressed gene candidate. In addition we tested also five other gene candidates (transcription factors) reported to be responsible for pod dehiscence in other plant species (bHLH Basic helix-loop-helix proteins *INDEHISCENT*, *SPATULA*, *SHATTERPROOF*, Basic Leucine Zipper Domain genes *bZIP* and *SHATTERING*; Ferrándiz et al., 2000; Girin et al., 2011; Dong et al., 2014). For these experiments we used RNA from pod suture tissue of parental line JI92 (domesticated, indehiscent pod) and JI64 (wild, dehiscent pod) as well as of eight contrasting RILs. As a result

we detected over expression in indehiscent lines of two gene candidates (*SHATTERING* and *SHATTERPROOF*; Figure 10C) and the rest of tested genes (*INDEHISCENT*, *SPATULA*, *bZIP*) did not show differential expression. In the second experiment we tested gene expression of 20 gene candidates derived from MACE analysis. In this case, we tested parental lines JI64 and JI92 only. We used RNA from dorsal and ventral pod suture tissue of three developmental stages (10, 15, and 20 days after flowering). In this case we detected three genes (MACE-P004, MACE-P013, MACE-P015) corresponding to the down/over expression as in MACE results (Figure 10A). In *M. truncatula* genome homologs genes are: MACE-P004 transmembrane protein, putative (Medtr3g016200); MACE-P013 NADP-dependent malate dehydrogenase (Medtr1g090730), and MACE-P015 peptidoglycan-binding domain protein (Medtr2g079050). In one case (MACE-P009) qRT-PCR showed the opposite results than in MACE (Figure 10A). *Medicago* homolog of MACE-P009 gene is cathepsin B-like cysteine protease (Medtr7g111060). The rest of candidate genes for pod dehiscence generated by MACE were tested by qRT-PCR, namely: glycosyltransferase family 92 protein (MACE-P001, Medtr2g437660); serine carboxypeptidase-like protein (MACE-P002, Medtr3g434850); KDEL-tailed cysteine endopeptidase CEP1 (MACE-P003, Medtr3g075390); 60S ribosomal protein L3B (MACE-P005, Medtr1g098540); dormancy/auxin associated protein

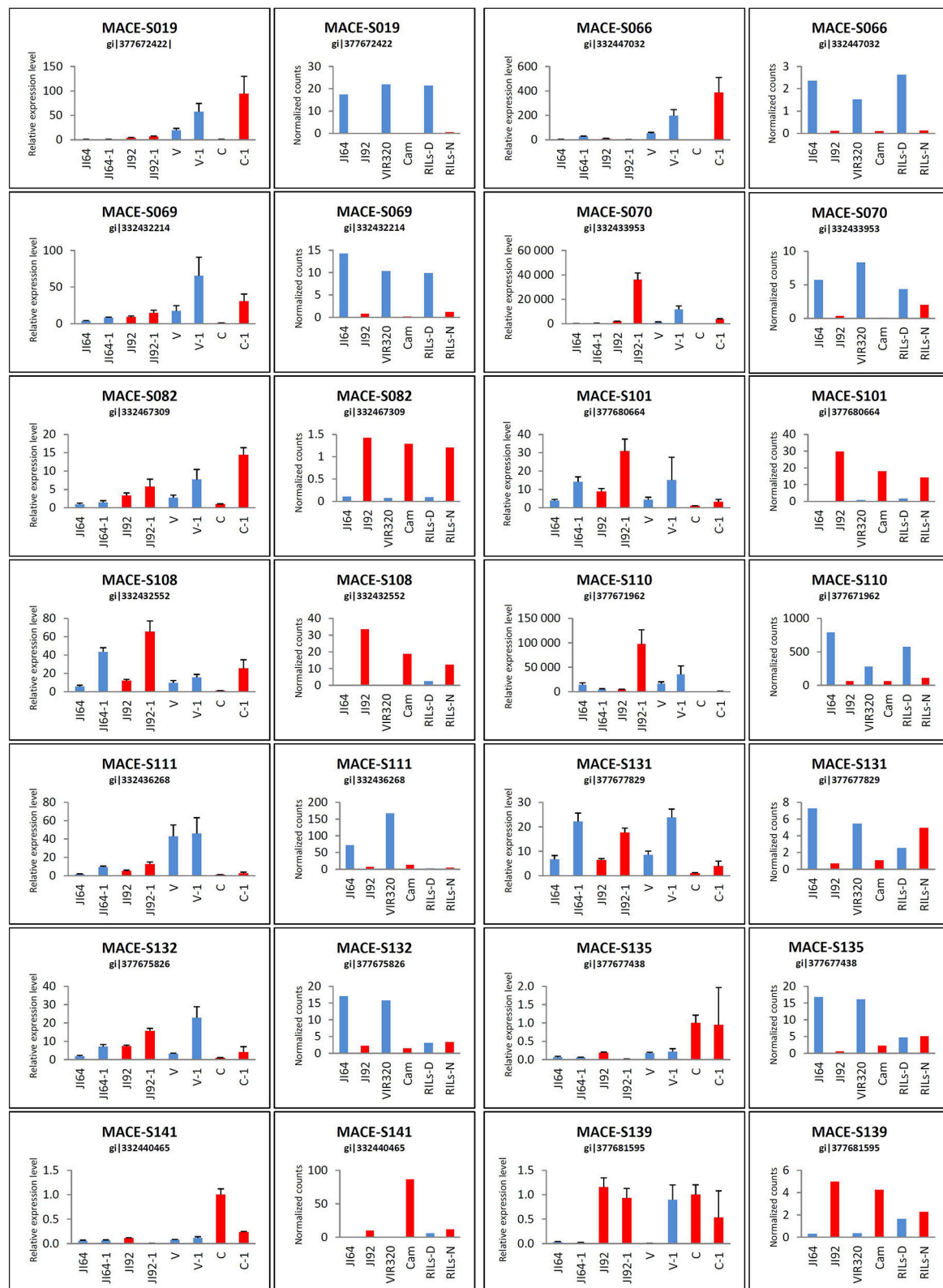


FIGURE 9 | qRT-PCR results of selected candidate genes in comparison with MACE analysis. qRT-PCR for four genotypes in two stages: JI64, JI92, V (VIR320), and C (Cameo)—younger stage; JI64-1, JI92-1, V-1, C-1—older stage and MACE for JI64, JI92, VIR320, Cam, RILs-D (dormant), and RILs-N (non-dormant).

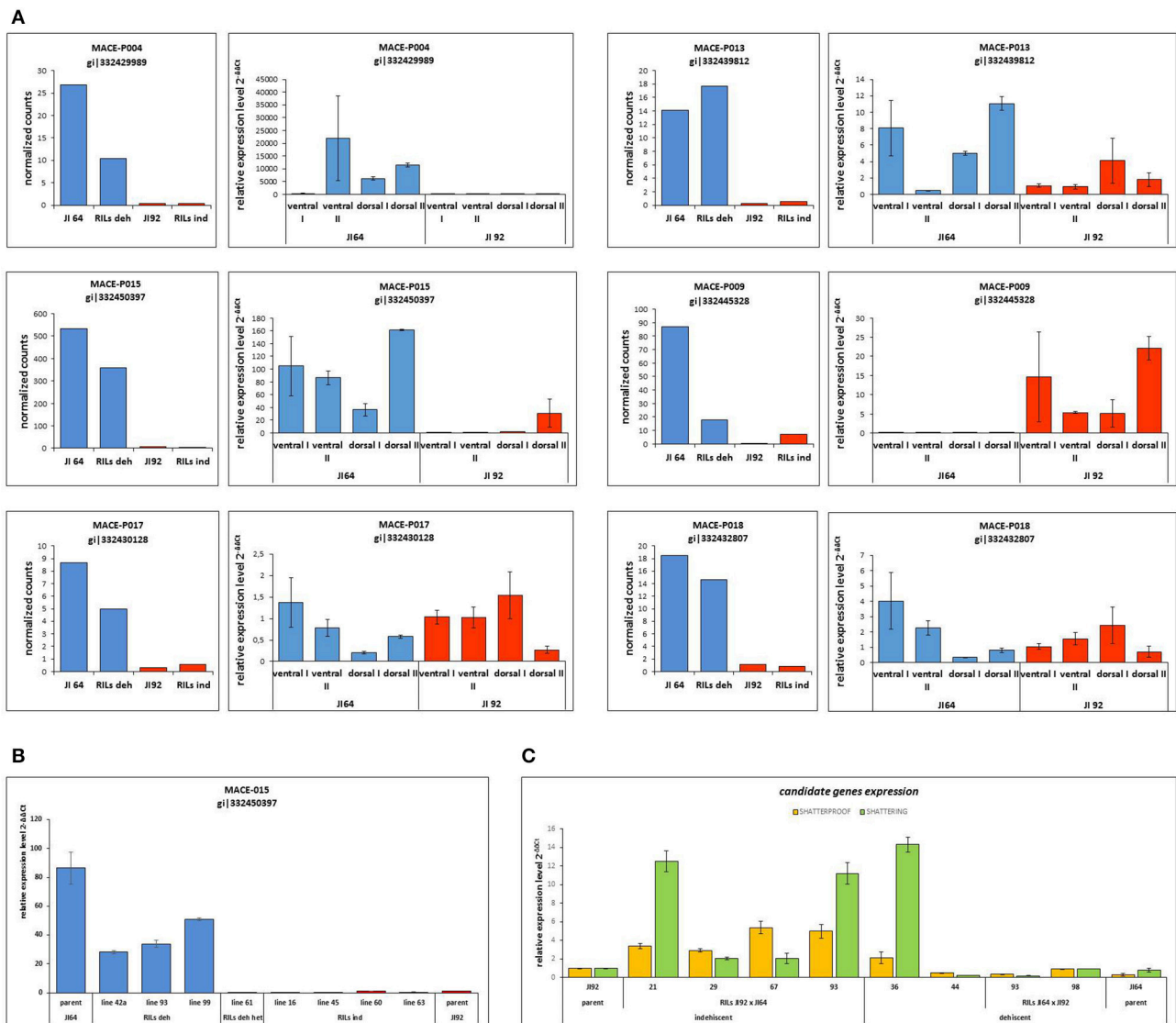


FIGURE 10 | qRT-PCR validation of selected DEGs detected to be associated with pod dehiscence. **(A)** MACE results in comparison with qRT-PCR of selected candidate genes. MACE-P004, MACE-P013, and MACE-P015 shows the same trend of over/down expression between MACE results of parental lines and RILs (JI 64—dehiscent pod, bulk of dehiscent RILs, JI 92—indehiscent pod, bulk of indehiscent RILs) and qRT-PCR of cDNA of parental lines (ventral or dorsal pod suture in two stages: I—younger stage, II—older stage). MACE-P009 showed the opposite results and MACE-P017 and MACE-P018 without any trend of down or over expression in qRT-PCR. **(B)** qRT-PCR results of contrasting RILs (dehiscent/indehiscent pod) in comparison with parental lines of candidate gene MACE-P015. **(C)** qRT-PCR results of contrasting RILs (dehiscent/indehiscent pod) in comparison with parental lines of candidate genes SHATTERPROOF and SHATTERING.

(MACE-P006, Medtr7g112860); enoyl-CoA hydratase 2, peroxisomal protein (MACE-P007, Medtr3g115040); disease-resistance response protein (MACE-P008, Medtr2g035150); GASA/GAST/Snakin (MACE-P010, Medtr1g025220); dormancy/auxin associated protein (MACE-P011, Medtr7g112860); TCP family transcription factor (MACE-P012, Medtr2g090960); LL-diaminopimelate aminotransferase (MACE-P014, Medtr2g008430); zinc finger A20 and AN1 domain stress-associated protein (MACE-P016, Medtr2g098160); auxin-responsive AUX/IAA family protein (MACE-P017, Medtr1g080860); huntingtin-interacting K-like

protein (MACE-P018, Medtr2g034010); transmembrane-like protein (MACE-P019, Medtr2g038550) and vacuolar processing enzyme (MACE-P020, Medtr4g101730). In the third experiment, we tested only candidates which showed differences in expression in both previous experiments using eight RILs with clearly determined phenotype (dehiscent or indehiscent pods). Four RILs were phenotypically dehiscent and four indehiscent. In this case, we used cDNA derived from the mixture of dorsal and ventral pod sutures tissue (Figure 11) for analysis. Finally, we tested four genes: MACE-P004 (homolog of transmembrane protein gene in *M. truncatula*, *P. sativum*

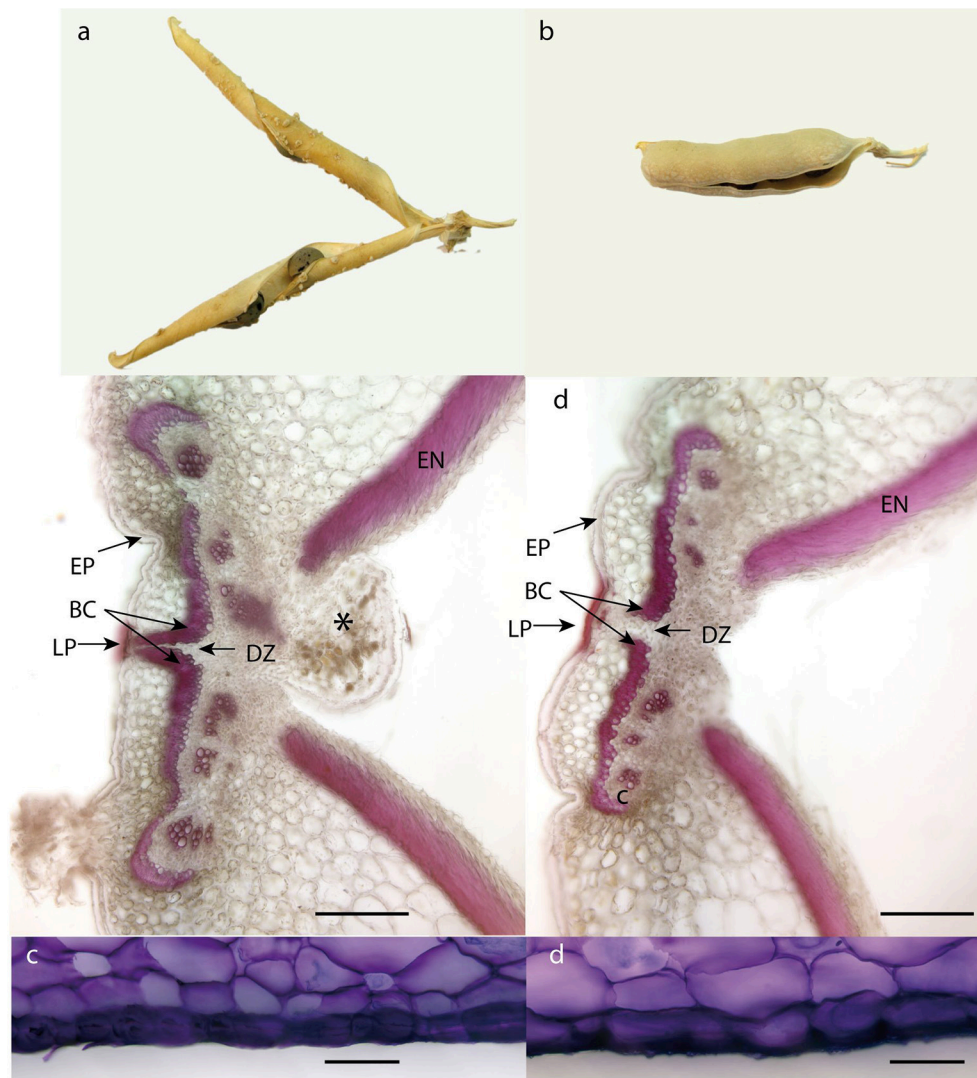


FIGURE 11 | Dorsal side of pea dehiscent and indehiscent pod. (a) Mature pod of dehiscent line (JI 64). **(b)** Mature pod of indehiscent line (JI 92). **(c,d)** Sections of dorsal side of dehiscent and indehiscent pea pod stained with fluoroglucinol: *, funikulus; BC, bundle cap; DZ, dehiscence zone; EN, endocarp, including inner sclerenchyma; EX, exocarp; EP, epidermis; DZ, dehiscence zone; LP, lignified epidermal plate.

LGIII: 24.5); MACE-P009 (homolog of cathepsin B-like cysteine protease in *M. truncatula*, *P. sativum* LGV: 20.3); MACE-P015 (homolog of peptidoglycan-binding domain protein gene in *M. truncatula*, *P. sativum* LGIII: 103.2); and Shatterproof (homolog of MADS-box transcription factor in *M. truncatula*, *P. sativum* LGIII: 89.3). As a result we didn't find any trend in down or over expression in contrasting RILs in relationship to dehiscence levels, with only one exception of MACE-P015 (homolog of peptidoglycan-binding domain protein gene in *M. truncatula*, *P. sativum* LGIII: 103.2) which showed trend of down expression in all tested indehiscent RILs (**Figure 10B**) with only one exception—RIL 61 (JI 64 × JI 92). Base on screening of PCR length polymorphism we recognized this line 61 (F_6) as heterozygous in MACE-P015 candidate gene because of presence of both parental alleles.

GO Annotation, KEGG

To annotate the DEGs in dehiscence-parents and RILs, the consensus MACE sequences were searched against the NCBI non-redundant protein database by blastx using *e*-value cut off of $1e^{-05}$. In biological process, most DEGs were found to be involved in metabolic processes (37 genes, 34%) which includes cellular metabolic process (28 genes, 26%) and primary metabolic process (28 genes, 26%), pigmentation (8 genes, 7%), regulation of biological process (8 genes, 7%), developmental processes such as anatomical structure development (3 genes, 3%; **Table S6**). In molecular function, the maximum DEGs were found to be involved in catalytic activity which includes mainly hydrolase activity (12 genes, 11%), transferase activity (8 genes, 7%) and oxidoreductase activity (7 genes, 6%). Next, the DEGs were mostly found to be involved in binding related activities such

as nucleic acid binding (10 genes, 9%) followed by nucleotide binding (9 genes, 8%) and nucleoside binding (5 genes, 5%). In the cellular component, the bulk of the DEGs belonged to cell part (41 genes, 38%) followed by membrane related proteins (22 genes, 20%), intracellular organelle (23 genes, 21%), and membrane-bound organelle (19 genes, 18%; **Table S6**). The KEGG pathway analysis showed that PsCam046431 is involved in phenylpropanoid biosynthesis, PsCam021037 in pentose phosphate, PsCam038465 in glycerophospholipid metabolism and PsCam042882 in carbon fixation pathways.

Genetic Mapping of Dehiscence Locus

We assume that dehiscent and indehiscent RILs bulks are contrasting for genomic regions which are responsible for dehiscence trait, because of selection for this trait and repeated selfing of RILs lines. On the other hand remaining genomic regions should be represented by polymorphic reads from RILs libraries due to mixing of RILs lines during bulking. When we compared polymorphism in reads from indehiscent RILs bulk and dehiscent RILs bulk three homozygous SNP rich genomic region associated with dehiscent trait were identified due to identification of homozygous SNP which indicated that this region was under selection during RILs lines development (**Figure S6**). Based on sequence homology first of them is located in the second half of *M. truncatula* chromosome 1. Second and third region are more clearer in contrast with first region. Second region is located at the beginning of chromosome 2 where homozygous SNP are concentrated around 2 megabase and third region is situated around 53 megabase at the end of chromosome 3. Others homozygous SNP are spread across all *Medicago* chromosomes and do not form distinct cluster.

DISCUSSION

Plant domestication process is interesting phenomenon of accelerated human directed evolution. To dissect genetic changes associated with this process, either wild to cultivated crosses and linkage mapping or newly genome wide association mapping are employed to infer on number of genes governing domestication traits. Although some of the genes underlying domestication traits were shown to be regulated at transcriptional level (Doebley et al., 1997; Konishi et al., 2006) limited studies were conducted to investigate transcriptomic changes between wild progenitors and cultivated crops, analyzing pod or seed tissues, such as wheat glumes (Zou et al., 2015). We used comprehensive transcriptomic, metabolomics, and anatomical analyses to compare domesticated and wild pea seed coats and pods in relation to the loss of seed dormancy and pod dehiscence.

Seed Coat Anatomical Structure and Histochemical Properties

Histological analysis of the seed coat in *M. truncatula* revealed changes in cell wall thickness in the outer integuments throughout seed development (Verdier et al., 2013a). In *Arabidopsis* and *Melilotus* (legume), seed permeability was modulated by mutations affecting extracellular lipid biosynthesis (in Verdier et al., 2013a). Similarly, in *M. truncatula*, cells

of the outer integument showed abundant accumulation of polyphenolic compounds (**Figure 4**); which upon oxidation may impact seed permeability (Moïse et al., 2005). Current knowledge about physical dormancy mainly comes from studies on morphological structure, phenolic content, and cuticle composition in legume species (reviewed in Smýkal et al., 2014). Morphological observation indicated that seed hardness was associated with the structure of palisade and cuticular layer (Vu et al., 2014) and presence or absence of cracks (Meyer et al., 2007; Koizumi et al., 2008). Other authors have proposed that the compositions of carbohydrates, hydroxylated fatty acids, or phenol compounds in seed coats control the level of permeability (Mullin and Xu, 2000, 2001; Shao et al., 2007; Zhou et al., 2010). Mullin and Xu (2001) found that the seed coat of an impermeable genotype had a high concentration of hemicellulose, essentially composed of xylans, which would reduce the hydrophilicity of the seed coat. We have found considerable structural and functional differences in testa properties between wild and domesticated peas. Contrary to Ma et al. (2004), who found small cuticular cracks in soft but not hard seeds of soybean, the surface was similar among used lines with only small discontinuities over the whole surface (**Figure 3**). However, we cannot exclude that these small fissures resulted from the SEM sample preparation, similarly to the above mentioned work. In the non-dormant genotypes subjected to imbibition, large fissures appear preferentially in the hilar region and strophiole (not shown). However, those are the most likely consequence of embryo imbibition and its volume increase and thus we do not expect those as primary sites of water entrance. There is a lack of detailed description of the primary pathway of water entry in pea as well as in the whole *Fabaceae* family, although the topic is thoroughly discussed (e.g., Baskin et al., 2000; Meyer et al., 2007; Ranathunge et al., 2010; Smýkal et al., 2014). It is thus not clear whether the hilar or strophiole region is the primary entrance of water in the non-dormant genotypes as suggested also by McDonald et al. (1988) or Korban et al. (1981) in soybean and common bean (Agbo et al., 1987) or if the minor fissures present in the cuticle and properties of the outer part of palisade macrosclereids make the difference as suggested by Ma et al. (2004). Interesting and generally neglected feature of macrosclereids is a presence of autofluorescent, phenolics containing lipidic material in the terminal caps of macrosclereids above the light line, which is not directly connected with the cuticle (**Figure S2**). There is no detailed information on the nature of this material or its possible functional significance. Obviously, detailed structure of outer part of the palisade macrosclereids deserves future attention. Dormant genotypes have thicker macrosclereids palisade layer, which might contribute to the water impermeability of coats of dormant pea genotypes as suggested by Miao et al. (2001). However, thickness alone does not necessarily account for water impermeability (de Souza and Marcos-Filho, 2001). Our results also suggest different thickness of palisade layer among the dormant genotypes with the most dormant J164 having the thickest palisade layer (**Figure 2**). Metachromatic staining with toluidine blue revealed that the non-dormant non-pigmented genotype Cameor exhibits high level of polyanionic pectins

with exposed free carboxyl groups (**Figure S3**). On the other hand, the non-dormant, but well-pigmented JI92 showed lower abundance of pectins related anions, similarly to pigmented dormant genotypes. There is clear connection between toluidine blue stainability and seed coat pigmentation—anthocyanidin presence. Taken together results from toluidine and Alcian blue with vaniline and DMACA suggest that tannins in the seed coats of pigmented peas are probably bound to other compounds of the cell walls, changing the staining properties of cell walls. Nature of these linkages is unknown, but covalently bound tannins might be indicated as alkaline hydrolysis releases proanthocyanidins from the cell walls (Krygier et al., 1982). Such expectation might be further supported by presence quercetin-3-rhamnoside fragments from LC/MS and MALDI-MS study. The detected increase in metachromatic staining of testa cell walls after weak acid treatment might be indicative of a crosslinking of pectins with other compounds in dormant as well as JI92 genotype, which is released in acid environment. Interestingly, it was reported that condensed tannins might be released from linkages in acid environment (Porter, 1989). We can speculate whether the intensity of pectin—tannin crosslinking is associated with physical dormancy of pea. There are scarce references indicating possible role of tannins in cell wall polymer network and its properties (Pizzi and Cameron, 1986). There is some supposition for callose deposition in the light line area (e.g., de Souza et al., 2012). However, strong staining with aniline blue fluorochrome staining in the upper part of macrosclereids including the light line (**Figure S2**) cannot be attributed to callose. Such assumption is consistent with the specific callose antibody localization which led us to the conclusion that the signal for aniline blue is probably signal for one or more other compounds structurally similar to callose. The interaction of aniline blue fluorochrome with other 1,3- or 1,4- β -D-glucans was described by Evans et al. and this interaction depends on the degree of polymerization, and nature of substitution of the 1,3- β -D-glucan chain as well as on the concentration of phosphate in the staining solution (Evans et al., 1984). Deeper anatomical and histochemical analysis of seed coat and more reliable detection of primary entrance point of water during early rehydration phase is needed.

Differentially Expressed Genes during Seed Coat Development and Seed Dormancy

Seed development has been thoroughly studied in number of crops including legumes, especially with focus on embryo development (Bewley et al., 2013). In our study, we have made comparative transcriptomics analysis in order to dissect candidate genes/pathways associated with domestication imposed changes on seed coat properties. Temporal transcriptional changes during seed and pod development were studied in pigeonpea (Pazhamala et al., 2016), soybean (Aghamirzaie et al., 2015; Redekar et al., 2015), *Medicago* (Gallardo et al., 2007; Benedito et al., 2008; Verdier et al., 2013a; Righetti et al., 2015), peanut (Zhu et al., 2014; Wan et al., 2016), and pea (Liu et al., 2015) seeds but no study was made on comparison of wild progenitor and cultivated crop. Moreover, these studies analyzed either entire seed/pod or developing embryos, while in our work we have used excised seed coat or dissected pod suture. During the RNA isolation from the seed

coat tissue, we have experienced great difficulties when working with wild pea seed samples. As reported earlier for pigmented soybean seeds (Wang and Vodkin, 1994) proanthocyanidins binds to RNA and prevent its extraction. We failed when using standard phenol/chloroform (McCarty, 1986) or guanidium thiocyanate (Chomczynski and Sacchi, 1987) methods, as well as common plant tissue RNA isolation kits.

There is problem of ambiguously mapped reads, which are the major source of error in RNA-Seq quantification (Robert and Watson, 2015). Short-read alignment is a complex problem due to the common occurrence of gene families. In contrast to RNA-seq, MACE methodology is derived from 3' UTR end of transcript and each is represented by single molecule. The choice of quantification tool also has large effect, as these also differ in the way they handle aligned data and multi-mapped/ambiguous reads (Robert and Watson, 2015). In order to exclude or minimize that identify DEGs are solely due to the genetic differences between contrasting parental genetic stocks we have used phenotypically classified bulks derived from RILs (**Table S1**, **Figure S1**). The concept of Bulk segregant analysis (BSA) was established as a method to detect markers in a specific genomic region by comparing two pooled DNA samples of individuals from a segregating population (Michelmore et al., 1991). Coupling BSA with the high throughput RNA sequencing has been shown to be an efficient tool for gene mapping and identification of differentially expressed genes (Chayut et al., 2015; Bojahr et al., 2016). One possible bottleneck of our analysis was bulking of several developmental stages into single MACE sample, where temporal and spatial expression might be hidden, resulting in differences between MACE and qRT-PCR data. Dynamic nature of gene expression both in spatial and temporal levels is clearly seen at qRT-PCR analysis of selected DEGs (**Figures 9**, **10**). The bulking of various developmental stages, moreover from different genotypes (in case of RILs) is the source of imprecision which can result in masking of DEGs. The key to the successful use of BSA is precision of phenotypic assignment. Although some imprecision in phenotypic classification and comparable low number of RILs used for bulking (**Table S1**), transcriptomics analysis has provided valid results. As shown on heat-map (**Figure 8**), RIL bulks are indeed genetic mixture of parental genotypes. MACE method (Kahl et al., 2012) detects allele-specific SNPs and indels associated with the defined genotypes that can be instantly used in genetic mapping (Bojahr et al., 2016). As result, there was significant clustering of homozygous SNPs associated with seed dormancy (e.g., respective parental alleles) on *Mt* chromosomes 3 and 4, or chickpea chromosomes 5 and 7, respectively (not shown). These correspond to pea linkage groups (LG) III and IV. Using identical RIL mapping population and DARTseq markers, we mapped seed coat thickness to LG I, III, IV, and VI (unpublished) and percentage of seed germination to LGII. These indicate that there is likely more than single major gene involved in seed dormancy, acting at different stages (testa thickness, permeability).

Despite that detected DEGs between dormant and non-dormant pea seeds belongs to various GO and KEGG pathways, the largest number of annotated ones was found within phenylpropanoid and flavonoid pathways (**Figure S4**). These

are involved in various activities such as UV filtration, fixing atmospheric nitrogen, and protection of cell walls (Zhao et al., 2013). Analysis of soybean mutant defective in seed coat led to identification of differentially expressed proline-rich and other cell wall protein transcripts (Kour et al., 2014). Moreover, this single gene mutation has resulted in differential expression of 1,300 genes, pointing out that complex series of events, many manifested at the transcript level, lead to changes in physiology, and ultimately structure of the cell wall. Similarly, we speculate that gene/-s causative of pea seed dormancy results in complex transcriptional and metabolomics changes. Recently, KNOTTED-like homeobox (KNOXII) gene, KNOX4, was found responsible for the loss of physical dormancy in the mutant *Medicago* seeds (Chai et al., 2016) resulting in differences in lipid monomer composition. These findings are in agreement with our data obtained by laser desorption-ionization mass spectrometric comparative analysis between dormant and non-dormant pea genotypes. Especially long chain hydroxylated fatty acids such as mono and dihydroxylated hexacosanoate, heptacosanoate, and octacosanoate were found in higher concentration in dormant peas compared to non-dormant ones (as shown in **Figure 8** for dihydroxyheptacosanoate and dihydroxyoctacosanoate), implying that the presence of a greater proportion of hydroxylated fatty acids may provide a greater interconnectivity of cutin hydrophobic components improving its stability and impermeability for water as discussed also by Shao et al. (2007). As downstream targets of KNOX4 gene, several key genes related to cuticle biosynthesis were identified, such as the cytochrome P450-dependent fatty acid omega-hydroxylase and fatty acid elongase 3-ketoacyl-CoA synthase (Chai et al., 2016). We have not found homologs genes when searched within our DEGs set. It can be hypothesized that different genes have been altered in independently domesticated crops, although possibly acting on identical pathways. There are limited studies combining of transcriptomic and metabolomics analysis (Enfissi et al., 2010) or a combination of both techniques including proteomics (Barros et al., 2010; Collakova et al., 2013). Such integrative approach enables not just to identify transcript and metabolite changes associated with given process, but also to focus on biochemical pathways relevant to studied trait and possibly also delimit candidate genes. As shown in *Medicago* (Verdier et al., 2013a) and soybean (Ranathunge et al., 2010) the gene expression in seed coat is complex and dynamic. Since we could not currently annotate 10% of detected transcripts, it can be expected that with available pea genome this could be further improved. They might include putative target candidate/-s for seed coat permeability.

Dormant Pea Seed Coat Accumulates More Proanthocyanidins

In legume seeds, there are three parts: the seed coat, the cotyledon, and the embryonic axis which, on average, represent 10, 89, and 1%, respectively, of the seed content. Seed coat pigmentation was shown to correlate with imbibition ability in several legumes, including common bean (Caldas and Blair, 2009), chickpea (Legesse and Powell, 1996), yardlong bean

(Kongjaimun et al., 2012), faba bean (Ramsay, 1997), and pea (Marbach and Mayer, 1974; Werker et al., 1979). The presence of proanthocyanidins (PAs) in seed coats can be assessed by the appearance of brownish coloration, which is the result of PA oxidation by polyphenol oxidase (Marles et al., 2008). In soybean (*Glycine max*), the recessive *i* allele results in high anthocyanin accumulation in the seed coat, resulting in dark brown or even black color (Tuteja et al., 2004; Yang et al., 2010). In contrast, the dominant *I* allele, which silences chalcone synthase (CHS) expression and hence blocks both anthocyanin and PA biosynthesis, results in a completely colorless seed coat. However, there is not simple relationship between the testa pigmentation imposing dormancy, as numerous cultivated pea varieties have colored testa yet do not display seed dormancy as illustrated by this study used JI92 landrace. Mendel's *A* gene beside flower color has pleiotropic effect including seed coat pigmentation (Hellens et al., 2010), yet these traits can be decoupled by recombination (Smykal, unpublished). Second, *B* gene of pea encodes a defective flavonoid 3', 5'-hydroxylase, and confers pink flower color, by control of hydroxylation of flavonoid precursors (Moreau et al., 2012). Neither this mutation results in alteration of seed dormancy. Comparably more is known on *Arabidopsis* and *Medicago* seed development owing to available mutants. Many of these mutations indicate the important role of proanthocyanidins and flavonoid pigments in testa development (Graeber et al., 2012) including effect on seed dormancy. The proanthocyanidins (PAs) received particular attention due to their abundance in seed coats (Dixon et al., 2005; Zhao et al., 2010) including pea (Ferraro et al., 2014). PAs are also known as the chemical basis for tannins, which are considered to be important part of physical dormancy in some species (Kantar et al., 1996; Ramsay, 1997). Flavan-3-ol-derived PA oligomers and anthocyanins are derived from the same precursors, proanthocyanidins (Lepiniec et al., 2006) and chemical diversity is introduced early in the pathway by cytochrome P450 enzymes (reviewed in Li et al., 2016a). Anthocyanidin synthase, anthocyanidin reductase, and leucoanthocyanidin reductase were studied at transcriptional level by Ferraro et al. (2014) in cultivated pea varieties and showed to be developmentally regulated. In our comparative transcriptome profiling we have not found any of these genes to be among DEGs, suggesting that differences in metabolites (quercetin, gallicocatechin) found by chemical analysis are not at these steps of PA biosynthesis. Anthocyanidins are either immediately modified by glycosylation to give anthocyanins by anthocyanidin 3-O-glycosyltransferases (UGTs) or reduced to generate flavan-3-ols (such as epicatechin) by anthocyanidin reductase for PA biosynthesis (Xie et al., 2003, 2004). There are several described *Medicago* mutants defective in respective genes, resulting in reduced testa pigmentation (Li et al., 2016b), although the relationship to seed dormancy was not specifically investigated. Indeed, we have detected several differentially expressed UDP-glycosylases, two of them studied by qRT-PCR (**Figure 9**). The glycosyltransferase superfamily consists of 98 subfamilies and only few have been characterized so far. Only a few members of the UGT72 family been shown to have activity toward flavonoids; such as the seed coat-specific UGT72L1 from *Medicago* (Zhao et al., 2010) and several seed specific

UGTs in *L. japonicus* (Yin et al., 2017). The UGT72L1 catalyzes (Zhao et al., 2010) formation of epicatechin 3'-O-glucoside (E3'OG), the preferred substrate for MATE transporters. MATE1 mutant display altered seed coat structure and PAs accumulation (Zhao and Dixon, 2011) and also has significantly lower seed dormancy levels (Smýkal, unpublished). The mechanism of PA polymerization is still unclear, but may involve the laccase-like polyphenol oxidase (Zhao et al., 2010). Notably, *Medicago myb5* and *myb14* mutants exhibit darker seed coat color than wild-type plants, with *myb5* also showing deficiency in mucilage biosynthesis, and accumulating only of the PA content of wild-type plants. When *myb5* seeds are exposed to water, they germinate readily without dormancy typical of wild type *Medicago* seeds (R. Dixon, personal communication and P. Smýkal, unpublished). All these observations suggest that PA oligomers play indeed a role in seed coat mediated dormancy.

Our LC/ESI-MS experiments confirm the presence of significantly higher contents of dimer and trimer of galocatechin (i.e., soluble tannins of prodelphinidin type) in dormant compared to non-dormant pea genotypes. Catechin dimer and trimer was also found in the pea seed coat extracts but their differences between dormant and non-dormant peas are much less significant than their galocatechin counterparts. This fact points out the significance of hydroxylation of B-ring of PAs in relation to dormancy. The insoluble PAs are the result of oxidative cross-linking with other cell components. Variation in PA content in the pea seeds has been reported (Troszyńska and Ciska, 2002) but not in comparison of wild vs. cultivated peas. PAs play also important roles in defense to pathogens, and because of the health benefits are of industry and medicine interest. PA biosynthesis and its regulation have been dissected in *Arabidopsis* using *transparent testa* (*tt*) mutants, which regulate production, transport or storage of PAs (Lepiniec et al., 2006), and 20 genes affecting flavonoid metabolism were characterized at the molecular level (reviewed in Bradford and Nonogaki, 2009). Many of these flavonoid biosynthesis pathway genes have been found to affect dormancy of *Arabidopsis* seeds, indicating the role of pigments in this process (Debeaujon et al., 2000). Similarly *Medicago* also synthesizes PAs in the seed coat, which consists essentially of epicatechin units (Lepiniec et al., 2006; Zhao et al., 2010). Polymerization of soluble phenolics to insoluble polymers is promoted by peroxidases (Gillikin and Graham, 1991) and catecholoxidases (Marbach and Mayer, 1974; Werker et al., 1979), which are abundant in legume seed coats. Positive correlation in content of phenolics, the requirement of oxidation and the activity of catechol oxidase in relation to seed dormancy (germination) in wild vs. domesticated pea seeds have been shown by Marbach and Mayer (1974) and Werker et al. (1979). Recently, epicatechin, cyanidin 3-O-glucoside, and delphinidin 3-O-glucoside were isolated in wild compared to cultivated soybean seed coats (Zhou et al., 2010) with epicatechin being in significant positive correlation with hardseededness.

Beside proanthocyanidin also flavonols, first of all quercetin derivatives, are frequently found in legumes including pea (Dueñas et al., 2004). We found significantly higher content of quercetin rhamnoside in dormant JI64 genotype compared

to non-dormant JI92. Similarly, its hydroxylated analog, i.e., myricetin-3-rhamnoside appeared to be a marker of dormancy. This compound was found in many legumes including chickpea, horse gram (Sreerama et al., 2010) and pea (Dueñas et al., 2004). As described in review of Agati et al. (2012) the antioxidant properties of flavonoids represent a robust biochemical trait of organisms exposed to oxidative stress of different origin during plant-environment interactions (regulation of the action of reaction oxygen species, ROS). The effect of ROS on the plant developmental processes including seed germination was described by Singh et al. (2016) including increase of free radical scavenging during pea seed germination (Lopez-Amoros et al., 2006). Presence of phenolic compounds in seed coat might help to protect against fungal diseases during germination as shown in lentil (Matus and Slinkard, 1993).

Pod Dehiscence

Pod maturation might terminate with pod shattering, which is an important trait for seed dispersal of wild species but generally unwanted trait in crops (Fuller and Allaby, 2009). Central to the ballistic seed dispersal in *Pisum* is the dehiscent pod (single carpel fused along its edges) where the central pod suture undergoes an explosive rupturing along a dehiscence zone (Ambrose and Ellis, 2008). During pod shattering, the two halves of the pod detach due to a combination of the diminished cell walls adhesion in the dehiscence zone, and the tensions established by the specific mechanical properties of drying cells of endo and exocarp of the pod shell. These two principal aspects are shared among families producing dry dehiscent fruit, such as *Fabaceae* and *Brassicaceae* (Grant, 1996; Dong and Wang, 2015). The spring-like tension within the pod shell is generated during differential drying-induced shrinkage of endocarp and the outer part of the shell—exocarp (Armon et al., 2011). Properties of both the obliquely arranged rigid and lignified inner sclerenchyma as well as the pectin rich exocarp cell with longitudinal orientation are crucial factors of springing in *Fabaceae* pods. Regulators of their development affecting geometrical arrangement of the layers and their histological properties (cell wall thickness and composition, lignification, hydration) will be key factors generating required tension. Composition and characteristics of pod cell shell cell walls correlate with shattering of yardlong bean and wild cowpea (Suanum et al., 2016) and thickness of the shell and extend of sclerenchymatous dorsal bundle caps was connected with shattering in soy (Tiwari and Bhatia, 1995). The major QTL controlling pod dehiscence in soybean is *qPDH1* (QTL for Pod Dehiscence 1). *qPDH1* had been recently cloned and shown to encode a dirigent-like protein expressed in the sclerenchyma of differentiating endocarp and modulating the mechanical properties of the pod shell. Lignin biosynthesis is the most likely process affected by *qPDH1* (Suzuki et al., 2010; Funatsuki et al., 2014), which might be connected with modulation of torsion within drying pod walls (Funatsuki et al., 2014). However, the precise biochemical activity of *qPDH1* is still unclear. Similarly the sclerenchyma differentiation and lignification of the endocarp and valve margin cells of *Arabidopsis* are central to silique dehiscence with *NAC SECONDARY WALL THICKENING PROMOTING FACTOR 1*

(*NST1*) and *SECONDARY WALL ASSOCIATED NAC DOMAIN PROTEIN 1* (*SDN1*) being identified as master regulators of their differentiation (Zhong et al., 2010). Lignification of endocarp and valve margin cells is lost together with dehiscence in the *nst1sdn1* double mutant (Mitsuda and Ohme-Takagi, 2008). The other decisive point of pod dehiscence is mechanical stability/instability of dehiscence (suture) zone which might trigger the explosive release of pod shell tension (Grant, 1996). Decreased cell to cell adhesion, might be carried out by action of endo-1,4-glucanases and endopolygalacturonases disintegrating the middle lamella in the separation layer (Christiansen et al., 2002). Degradation of pectin in the middle lamella of abscission zone is common theme in fruit shattering (Dong and Wang, 2015). Contrary, mechanism of pod shattering resistance due to reinforcement of the suture has been described in domesticated soybeans *NST1/2* homologous transcription factor *SHATTERING1-5* (*SHAT1-5*) from NAC family has been unveiled, inducing excessive secondary cell wall deposition and lignification in the outer part of the suture (fiber cap cells; Dong et al., 2014). Up-regulated expression of *SHAT1-5* in domesticated soy “locks” the dehiscence zone interconnecting the vascular bundle caps of sclerenchymatous fibers in the suture vicinity preventing shattering.

The fundamental elements of fruit shattering regulatory network is being uncovered recently in *Arabidopsis* and homologous genes were identified also in other species and crops (*Zea mays*, *Triticum aestivum*, *Oryza sativa*, *Glycine max*, *Sorghum bicolor*, *Sorghum propinquum*, or *Solanum lycopersicum*; for review see Dong and Wang, 2015; Ballester and Ferrandiz, 2016). MADS box genes of *Arabidopsis* *SHATTERPROOF1* (*SHP1*) and *SHATTERPROOF2* (*SHP2*) participate in the dehiscence zone specification (Liljegren et al., 2000). *INDEHISCENT* (*IND*) b-HLH transcription factor act down-stream to *SHP1/2* as regulator sclerenchyma differentiation in endocarp and valve margin. The *shp1/2* double mutant as well as *ind* produces indehiscent siliques devoid of proper cell specification and differentiation in the dehiscence zone (Liljegren et al., 2004; Dong and Wang, 2015). *SHP1* and *SHP2* are required for the proper specification of the different cell types within the valve margin and the DZ and both genes probably represent the top of the hierarchy regulating DZ formation (Liljegren et al., 2000). *SHATTERPROOF* genes along with *INDEHISCENT* (*IDEH*) are the main regulators of establishment of lignified layer, which causes pod dehiscence. Another MADS box gene involved in dehiscence zone formation is *FRUITFULL* (*FUL*), which expression appears at the inception of the carpel primordia, and soon after becomes restricted to the cells that will give rise to the valves, a pattern that is complementary to that of the *SHP* genes (Liljegren et al., 2004; Dong and Wang, 2015). Significant up-regulation of *SHATTERING1-5* (*SHAT1-5*) in the fiber cap cells (FCC) of cultivated soybean was shown to be responsible for the excessive cell wall deposition in the FCC, which in turn prevents the pod from committing dehiscence after maturation (Dong et al., 2014). Homologous genes defining dehiscence zone identity and its differentiation *INDEHISCENT*, *SPATULA*, *SHATTERPROOF*, *bZIP*, and *SHATTERING* (Ferrándiz et al., 2000; Girin et al.,

2011; Dong et al., 2014) were identified in *Pisum* genome and their abundance was tested in RNA isolated from pod suture tissue of wild and domesticated pea as well as of contrast RILs incurred from crossing of wild and domesticate parent with dehiscent or indehiscent pods. In case of *SHATTERPROOF* and *SHATTERING* homologous genes, we found differences in expression between parental lines of pea but we didn't find the similar results in case of contrast RILs. The other homologous genes (*INDEHISCENT*, *SPATULA*, and *bZIP*) did not exhibit any significant difference in expression between dehiscent and non-dehiscent phenotypes.

In the legumes, the pod shattering trait is controlled by one or two dominant genes or QTL. In pea and lentil, genes controlling pod shattering map to a syntenic region, suggesting that the same genes may have been modified during the domestication of the two cool-season legumes (Weeden et al., 2002; Weeden, 2007). Single locus control of pod dehiscence was found in lentil (Ladizinsky, 1998), while two loci in mungbean (Isemura et al., 2012), yardlong bean (Kongjaimun et al., 2012), one controlling the number of twists along the length of the shattered pod, and second the percentage of shattered pods, similarly to two loci found in pea (Weeden et al., 2002; Weeden, 2007), and common bean (Koinange et al., 1996). Bordat et al. (2011) localized *Dpo* locus responsible for loss of pea pod dehiscence on LGIII. We obtained the similar result by the genome-wide DArTseq analysis. Based on comparison of our candidate gene position in *M. truncatula* genome and SNPs map (Tayeh et al., 2015) we localized our candidate genes in pea genome. In total of our 25 candidate genes 7 are localized on LGIII, 6 on LGVI, 4 on LGII, 3 on LGV, 2 on LGVII, 2 on LGIV, and 1 on LGI (not shown). MACE-P015, the main candidate gene possibly responsible for pod dehiscence localized on LGIII, is a homolog of peptidoglycan-binding domain protein (PGDB) of *M. truncatula* (Medtr2g079050). These proteins may have a general peptidoglycan binding function and this motif is found at the N or C terminus of a variety of enzymes involved in bacterial cell wall degradation. Many of the proteins having this domain are so far uncharacterized. Matrix metalloproteinases (MMP), which catalyze extracellular matrix degradation, have N-terminal domains that resemble PGDB (Seiki, 1999). On the other hand our candidate MACE-P015 has also 80% match with *Cicer arietinum* proline-rich extensin-like protein EPR1 (XM_004488673). Extensins are plant specific structural cell-wall proteins (Lamport et al., 2011); they can account for up to 20% of the dry weight of the cell wall and can significantly modulate mechanical cell wall properties through linkages to other cell wall component, which can play a role in pod dehiscence.

New Insights into Pea Seed and Pod Development in Relation to Domestication

Study of biochemical and molecular mechanisms underlying plant domestication process is important area of research in plant biology. In the current study we used a comparative anatomy, metabolomics, and transcriptome profiling of pods and seed coats in wild and domesticated pea in order to identify genes associated with loss of seed dormancy as well as pod dehiscence.

We have identified genes showing differential expression in respective parents as well as phenotypically contrasting RILs. Among others, there were number of genes belonging to phenylpropanoid pathway, which was also identified by metabolomics analysis of seed coat. Our results support the role of proanthocyanidins and their derivatives in physical seed coat mediated dormancy. One of the identified differentially expressed gene involved in pod dehiscence showed significant down-expression in dorsal and ventral pod suture of indehiscent genotypes. Moreover, this homolog of peptidoglycan-binding domain or proline-rich extensin-like protein mapped correctly to predicted *Dpo1* locus on PsLGIII. This integrated analysis of the seed coat in wild and cultivated pea raised new questions associated with domestication and seed dormancy. Having underlying gene(s) in hands for various independently domesticated legume crops it would help our understanding of genetic and molecular processes involved in seeds dormancy. Moreover, extended knowledge on control seed dispersal and seed dormancy is necessary for diverse applications—biodiversity conservation as well as breeding.

AUTHOR CONTRIBUTIONS

Conceived and designed experiments: PS, PB, AS, and PH. Performed experiments: IH, PS, AJ, PH, LP, MC, and MV. Analyzed the data: IH, AJ, PB, OT, and KA. Wrote the paper: IH, AJ, OT, PH, and PS. All authors have read and approved the manuscript.

ACKNOWLEDGMENTS

This research was funded by Grant Agency of Czech Republic, 14-11782S project. OT was supported by partial institutional funding on long-term conceptual development of Agricultural Research, Ltd. organization. LP was supported by partial institutional funding IGA FA MENDELU No. IP 22/2016. AS, AJ, PB, MC, and MV were supported by project Czech Ministry of Education, Youth and Sports—Project LO1417 (AS, AJ) and LO1305 (PB, MC, and MV).

REFERENCES

- Abbo, S., Rachamim, E., Zehavi, Y., Zekai, I., Lev-Yadun, S., and Gopher, A. (2011). Experimental growing of wild pea in Israel and its bearing on near Eastern plant domestication. *Ann. Bot.* 107, 1399–1404. doi: 10.1093/aob/mcr081
- Agati, G., Azzarello, E., Pollastri, S., and Tattini, M. (2012). Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci.* 196, 67–76. doi: 10.1016/j.plantsci.2012.07.014
- Agbo, G. N., Hosfield, M. A., Uebbersax, M. A., and Klomparens, K. (1987). Seed microstructure and its relationship to water uptake in isogenic lines and a cultivar of drybeans (*Phaseolus vulgaris* L.). *Food Microstruct.* 6, 91–102.
- Aghamirzaie, D., Batra, D., Heath, L. S., Schneider, A., Grene, R., and Collakova, E. (2015). Transcriptome-wide functional characterization reveals novel relationships among differentially expressed transcripts in developing soybean embryos. *BMC Genomics* 16:928. doi: 10.1186/s12864-015-2108-x
- Alves-Carvalho, S., Aubert, G., Carrere, S., Cruaud, C., Brochot, A.-L., Jacquin, F., et al. (2015). Full-length *de novo* assembly of RNAseq data in pea (*Pisum*

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00542/full#supplementary-material>

Table S1 | Germination characteristics of parental and RIL lines.

Table S2 | Summary statistics for MACE sequencing data and annotation to the *Pisum* mRNA sequences from NCBI and to the contigs based on the *de novo* assembly of sequences.

Table S3 | List of primers used for qRT-PCR.

Table S4 | Complete set of DEGs detected in seed dormancy and pod dehiscence experiments, between wild and cultivated parental genotypes as well as RILs, with indicated homology to annotated pea transcriptome (Tayeh et al., 2015) and MACE expression values.

Table S5 | GO annotations of seed dormancy genes.

Table S6 | GO annotations of pod dehiscence genes.

Figure S1 | Germination indexes of all 126 RIL lines tested at 25°C over the period of 7 days. Ordered by cumulative germination percentages (A) with shown Coefficient of velocity (B), Timson indexes (C), and Mean germination time (D) arranged accordingly.

Figure S2 | Seed coat transverse sections from extrahilar region: Cameor (a), JI92 (b), JI64 (c), and VIR320 (d). UV excited autofluorescence (a–d): white arrow, cuticle; asterisk, light line. Sudan Red 7B staining of terminal parts of macrosclereids (e–h): white arrow, cuticle; black arrow, lipidic material different from the cuticle. Aniline blue fluorochrome staining under UV excitation (i–l): asterisk, light line; white arrow, the edge between pigmented and non-pigmented interface of JI92, inlay in (f) callose immunodetection (blue excitation); scale bars = 25 μm

Figure S3 | Seed coat transverse sections from extrahilar region: Cameor (a), JI92 (b), JI64 (c), and VIR320 (d). Metachromatic toluidine blue staining is indicative of high density of polyanionic surface (a–d): asterisk, light line. Toluidine blue staining after mild acid treatment (e–h): black arrow = the edge between pigmented and non-pigmented interface of JI92; scale bar = 50 μm.

Figure S4 | KEGG phenylpropanoid (A) and flavonoid (B) pathways of DEGs between dormant and nondormant seeds.

Figure S5 | KEGG phenylpropanoid pathway of DEGs between dehiscent and indehiscent pods.

Figure S6 | Strictly homozygous SNP between dehiscence and indehiscence RIL bulks mapped to the eight *Medicago truncatula* chromosomes.

- sativum* L.) provides a gene expression atlas and gives insights in root nodulation in this species. *Plant J.* 8, 1–19. doi: 10.1111/tpj.12967
- Amarowicz, R., Estrella, I., Hernández, T., Dueñas, M., Troszyńska, A., Kosińska, A., et al. (2009). Antioxidant activity of a red lentil extract and its fractions. *Int. J. Mol. Sci.* 10, 5513–5527. doi: 10.3390/ijms10125513
- Ambrose, M. J., and Ellis, T. H. N. (2008). Ballistic seed dispersal and associated seed shadow in wild *Pisum* germplasm. *Pisum Genet.* 40, 5–10.
- Appelhaagen, I., Thiedig, K., Nordholt, N., Schmidt, N., Huep, G., Sagasser, M., et al. (2014). Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta* 240, 955–970. doi: 10.1007/s00425-014-2088-0
- Armon, S., Efrati, E., Kupferman, R., and Sharon, E. (2011). Geometry and mechanics in the opening of chiral seed pods. *Science* 333, 1726–1730. doi: 10.1126/science.1203874
- Bajaj, D., Das, S., Upadhyaya, H. D., Ranjan, R., Badoni, S., Kumar, V., et al. (2015). A genome-wide combinatorial strategy dissects complex genetic architecture of seed coat color in chickpea. *Front. Plant Sci.* 6:979. doi: 10.3389/fpls.2015.00979

- Ballester, P., and Ferrandiz, C. (2016). Shattering fruits: variations on a dehiscent theme. *Curr. Opin. Plant Biol.* 35, 68–75. doi: 10.1016/j.pbi.2016.11.008
- Barros, E., Lezar, S., Anttonen, M. J., van Dijk, J. P., Röhligh, R. M., Kok, E. J., et al. (2010). Comparison of two GM maize varieties with a near-isogenic non-GM variety using transcriptomics, proteomics and metabolomics. *Plant Biotechnol. J.* 8, 436–451. doi: 10.1111/j.1467-7652.2009.00487.x
- Baskin, J. M., and Baskin, C. C. (2004). A classification system for seed dormancy. *Seed Sci. Res.* 14, 1–16. doi: 10.1079/SSR2003150
- Baskin, J. M., Baskin, C. C., and Li, X. (2000). Taxonomy, anatomy and evolution of physical dormancy in seeds. *Plant Species Biol.* 15, 139–152. doi: 10.1046/j.1442-1984.2000.00034.x
- Beisson, F., Li, Y., Bonaventure, G., Pollard, M., and Ohlrogge, J. B. (2007). The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of *Arabidopsis*. *Plant Cell* 19, 351–368. doi: 10.1105/tpc.106.048033
- Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., et al. (2008). A gene expression atlas of the model legume *Medicago truncatula*. *Plant J.* 55, 504–513. doi: 10.1111/j.1365-3113.2008.03519.x
- Bennett, E. J., Roberts, J. A., and Wagstaff, C. (2011). The role of the pod in seed development: strategies for manipulating yield. *New Phytol.* 190, 838–853. doi: 10.1111/j.1469-8137.2011.03714.x
- Bewley, J. D. (1997). Seed germination and dormancy. *Plant Cell* 9, 1055–1066. doi: 10.1105/tpc.9.7.1055
- Bewley, J. D., Bradford, K., Hilhorst, H., and Nonogaki, H. (2013). *Seeds: Physiology of Development, Germination and Dormancy*, 3rd Edn. New York, NY: Springer-Verlag. doi: 10.1007/978-1-4614-4693-4
- Bogdanova, V. S., Galieva, E. R., Yadrinskiy, A. K., and Kosterin, O. E. (2012). Inheritance and genetic mapping of two nuclear genes involved in nuclear-cytoplasmic incompatibility in peas (*Pisum sativum* L.). *Theor. Appl. Genet.* 124, 1503–1512. doi: 10.1007/s00122-012-1804-z
- Bojahr, J., Nhengiwa, O., Krezdorn, N., Rotter, B., Saal, B., Ruge-Wehling, B., et al. (2016). Massive analysis of cDNA ends (MACE) reveals a co-segregating candidate gene for *LpPgl* stem rust resistance in perennial ryegrass (*Lolium perenne*). *Theor. Appl. Genet.* 129, 1915–1932. doi: 10.1007/s00122-016-2749-4
- Bordat, A., Savoie, V., Nicolas, M., Salse, J., Chauveau, A., Bourgeois, M., et al. (2011). Translational genomics in legumes allowed placing *in silico* 5460 unigenes on the pea functional map and identified candidate genes in *Pisum sativum* L. G3 1, 93–103. doi: 10.1534/g3.111.000349
- Bradford, K., and Nonogaki, H. (2009). *Seed Development, Dormancy and Germination*. Annual Plant Review, Vol. 27. Oxford: Blackwell.
- Caldas, G. V., and Blair, M. W. (2009). Inheritance of condensed tannin content and relationship with seed colour and pattern genes in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 119, 131–142. doi: 10.1007/s00122-009-1023-4
- Chai, M., Zhou, C., Molina, I., Fu, C., Nakashima, J., Li, G., et al. (2016). A class II KNOX gene, KNOX4, controls seed physical dormancy. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6997–7002. doi: 10.1073/pnas.1601256113
- Chayut, N., Yuan, H., Ohali, S., Meir, A., Yeselson, Y., Portnoy, V., et al. (2015). A bulk segregant transcriptome analysis reveals metabolic and cellular processes associated with Orange allelic variation and fruit β -carotene accumulation in melon fruit. *BMC Plant Biol.* 15:274. doi: 10.1186/s12870-015-0661-8
- Chen, H., Osuna, D., Colville, L., Lorenzo, O., Graeber, K., Küster, H., et al. (2013). Transcriptome-wide mapping of pea seed ageing reveals a pivotal role for genes related to oxidative stress and programmed cell death. *PLoS ONE* 8:e78471. doi: 10.1371/journal.pone.0078471
- Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156–159.
- Christiansen, L. C., Dal Degan, F., Ulvskov, P., and Borkhardt, B. (2002). Examination of the dehiscence zone in soybean pods and isolation of a dehiscence-related endopolygalacturonase gene. *Plant Cell Environ.* 25, 479–490. doi: 10.1046/j.1365-3040.2002.00839.x
- Collakova, E., Aghamirzaie, D., Fang, Y., Klumas, C., Tabataba, F., Kakumanu, A., et al. (2013). Metabolic and transcriptional reprogramming in developing soybean (*Glycine max*) embryos. *Metabolites* 3, 347–372. doi: 10.3390/metabo3020347
- de Souza, F. H. D., and Marcos-Filho, J. (2001). A seed coat as a modulator of seed-environment relationships in *Fabaceae*. *Rev. Bras. Bot.* 24, 365–375. doi: 10.1590/s0100-84042001000400002
- de Souza, T. V., Voltolini, C. H., Santos, M., and Silveira Paulilo, M. T. (2012). Water absorption and dormancy-breaking requirements of physically dormant seeds of *Schizolobium parahyba* (Fabaceae–Caesalpinioideae). *Seed Sci. Res.* 22, 169–176. doi: 10.1017/S0960258512000013
- Debeaujon, I., Léon-Kloosterziel, K. M., and Koornneef, M. (2000). Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiol.* 122, 403–414. doi: 10.1104/pp.122.2.403
- Dixon, R. A., Sharma, S. B., and Xie, D. (2005). Proanthocyanidins—a final frontier in flavonoid research? *New Phytol.* 165, 9–28. doi: 10.1111/j.1469-8137.2004.01217.x
- Doebley, J., Stec, A., and Hubbard, L. (1997). The evolution of apical dominance in maize. *Nature* 386, 485–488. doi: 10.1038/386485a0
- Dong, Y., and Wang, Y.-Z. (2015). Seed shattering: from models to crops. *Front. Plant Sci.* 6:476. doi: 10.3389/fpls.2015.00476
- Dong, Y., Yang, X., Liu, J., Wang, B. H., Liu, B. L., Wang, Y. Z., et al. (2014). Pod shattering resistance associated with domestication is mediated by a NAC gene in soybean. *Nat. Commun.* 5:3352. doi: 10.1038/ncomms4352
- Duarte, J., Riviere, N., Baranger, A., Aubert, G., Burstin, J., Cornet, L., et al. (2014). Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. *BMC Genomics* 15:126. doi: 10.1186/1471-2164-15-126
- Dueñas, M., Estrella, I., and Hernández, T. (2004). Occurrence of phenolic compounds in the seed coat and the cotyledon of peas (*Pisum sativum* L.). *Eur. Food Res. Technol.* 119, 116–123. doi: 10.1007/s00217-004-0938-x
- Enfissi, E. M., Barneche, F., Ahmed, I., Lichtlé, C., Gerrish, C., McQuinn, R. P., et al. (2010). Integrative transcript and metabolite analysis of nutritionally enhanced DE-ETIOLATED1 downregulated tomato fruit. *Plant Cell* 22, 1190–1215. doi: 10.1105/tpc.110.073866
- Evans, N. A., Hoyne, P. A., and Stone, B. A. (1984). Characteristics and specificity of the interaction of a fluorochrome from aniline blue (sirofluor) with polysaccharides. *Carbohydr. Polym.* 4, 215–230. doi: 10.1016/0144-8617(84)90012-2
- Ferrándiz, C., Liljegren, S. J., and Yanofsky, M. F. (2000). Negative regulation of the SHATTERPROOF genes by FRUITFULL during Arabidopsis fruit development. *Science* 289, 436–438. doi: 10.1126/science.289.5478.436
- Ferraro, K., Jin, A. L., Nguyen, T.-R., Reinecke, D. M., Ozga, J. A., and Ro, D.-K. (2014). Characterization of proanthocyanidin metabolism in pea (*Pisum sativum*) seeds. *BMC Plant Biol.* 14:238. doi: 10.1186/s12870-014-0238-y
- Finch-Savage, W. E., and Leubner-Metzger, G. (2006). Seed dormancy and the control of germination. *New Phytol.* 171, 501–523. doi: 10.1111/j.1469-8137.2006.01787.x
- Foyer, C. H., Lam, H. M., Nguyen, H. T., Siddique, K. H. M., Varshney, R. K., Colmer, T. D., et al. (2016). Neglecting legumes has compromised human health and sustainable food production. *Nature Plants* 2, 1–10. doi: 10.1038/nplants.2016.112
- Franssen, S. U., Shrestha, R. P., Bräutigam, A., Bornberg-Bauer, E., and Weber, A. P. (2011). Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics* 12:227. doi: 10.1186/1471-2164-12-227
- Fuller, D. Q., and Allaby, R. G. (2009). Seed dispersal and crop domestication: shattering, germination and seasonality in evolution under cultivation. *Ann. Plant Rev.* 38, 238–295. doi: 10.1002/9781444314557.ch7
- Fuller, D. Q., Denham, T., Arroyo-Kalin, M., Lucas, M., Stevens, C. J., Qin, L., et al. (2014). Convergent evolution and parallelism in plant domestication revealed by an expanding archaeological record. *Proc. Natl. Acad. Sci. U.S.A.* 111, 6147–6152. doi: 10.1073/pnas.1308937110
- Funatsuki, H., Suzuki, M., Hirose, A., Inaba, H., Yamada, T., Hajika, M., et al. (2014). Molecular basis of a shattering resistance boosting global dissemination of soybean. *Proc. Natl. Acad. Sci. U.S.A.* 111, 17797–17802. doi: 10.1073/pnas.1417282111
- Gallardo, K., Firnhaber, C., Zuber, H., Hélicher, D., Belghazi, M., Henry, C., et al. (2007). A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds. *Mol. Cell. Proteomics* 6, 2165–2179. doi: 10.1074/mcp.M700171-MCP200
- Gardner, R. O. (1975). Vanillin - hydrochloric acid as a histochemical test for tannin. *Stain Technol.* 50, 315–317. doi: 10.3109/10520297509117081
- Gillilkin, J. W., and Graham, J. S. (1991). Purification and developmental analysis of the major anionic peroxidases from the seed coat of *Glycine max*. *Plant Physiol.* 96, 214–220. doi: 10.1104/pp.96.1.214

- Girin, T., Paicu, T., Stephenson, P., Fuentes, S., Korner, E., O'Brien, M., et al. (2011). INDEHISCENT and SPATULA interact to specify carpel and valve margin tissue and thus promote seed dispersal in *Arabidopsis*. *Plant Cell* 23, 3641–3653. doi: 10.1105/tpc.111.090944
- Graeber, K., Nakabayashi, K., Miatton, E., Leubner-Metzger, G., and Soppe, W. J. J. (2012). Molecular mechanisms of seed dormancy. *Plant Cell Environ.* 35, 1769–1786. doi: 10.1111/j.1365-3040.2012.02542.x
- Grant, W. F. (1996). Seed pod shattering in the genus *Lotus* (Fabaceae): a synthesis of diverse evidence. *Can. J. Plant Sci.* 76, 447–456. doi: 10.4141/cjps96-079
- Hammer, K. (1984). Das Domestikationssyndrom. *Kulturpflanze* 32, 11–34. doi: 10.1007/BF02098682
- Hedley, C. L., Smith, C. M., Ambrose, M. J., Cook, S., and Wang, T. L. (1986). An analysis of seed development in *Pisum sativum*. The effect of the *r*-locus on the growth and development of the seed. *Ann. Bot.* 58, 371–379. doi: 10.1093/oxfordjournals.aob.a087215
- Hellens, R. P., Moreau, C., Lin-Wang, K., Schwinn, K. E., Thomson, S. J., Fiers, M. W., et al. (2010). Identification of Mendel's white flower character. *PLoS ONE* 5:e13230. doi: 10.1371/journal.pone.0013230
- Isemura, T., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE* 7:e41304. doi: 10.1371/journal.pone.0139666
- Jang, S. J., Sato, M., Sato, K., Jitsuyama, Y., Fujino, K., Mori, H., et al. (2015). A single-nucleotide polymorphism in an endo-1,4- β -glucanase gene controls seed coat permeability in soybean. *PLoS ONE* 10:e0128527. doi: 10.1371/journal.pone.0128527
- Kahl, G., Molina, C., Rotter, B., Jüngling, R., Frank, A., Krezdorn, N., et al. (2012). Reduced representation sequencing of plant stress transcriptomes. *J. Plant Biochem. Biotechnol.* 21, 119–127. doi: 10.1007/s13562-012-0129-y
- Kalendar, R., Lee, D., and Schulman, A. H. (2014). "FastPCR software for PCR, *in silico* PCR, and oligonucleotide assembly and analysis," in *DNA Cloning and Assembly Methods, Vol. 1116, Methods in Molecular Biology*, eds S. Valla and R. Lale (New York, NY: Springer), 271–302. doi: 10.1007/978-1-62703-764-8_18
- Kantar, F., Pilbeam, C. J., and Hebblethwaite, P. D. (1996). Effect of tannin content of faba bean (*Vicia faba*) seed on seed vigour, germination and field emergence. *Ann. Appl. Biol.* 128, 85–93. doi: 10.1111/j.1744-7348.1996.tb07092.x
- Kaur, S., Pembleton, L. W., Cogan, N. O., Savin, K. W., Leonforte, T., Paull, J., et al. (2012). Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* 13:104. doi: 10.1186/1471-2164-13-104
- Kislev, M. E., and Bar-Yosef, O. (1988). "The legumes: the earliest domesticated plants in the Near East?" *Curr. Anthropol.* 29, 175–179. doi: 10.1086/203623
- Koinange, E. M. K., Singh, S., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x
- Koizumi, M., Kaori, K., Seiichiro, I., Nobuaki, I., Shigehiro, N., and Hiromi, K. (2008). Role of seed coat in imbibing soybean seed observed by micro-magnetic resonance imaging. *Ann. Bot.* 102, 343–352. doi: 10.1093/aob/mcn095
- Kongjaimun, A., Kaga, A., Tomooka, N., Somta, P., Vaughan, D. A., and Srinives, P. (2012). The genetics of domestication of yardlongbean, *Vigna unguiculata* (L.) Walp. ssp. *unguiculata* cv.-gr. *sesquipedalis*. *Ann. Bot.* 109, 1185–1200. doi: 10.1093/aob/mcs048
- Konishi, S., Izawa, T., Lin, S. Y., Ebana, K., Fukuta, Y., Sasaki, T., et al. (2006). An SNP caused loss of seed shattering during rice domestication. *Science* 312, 1392–1396. doi: 10.1126/science.1126410
- Korban, S. S., Coyne, D. P., and Weihing, J. L. (1981). Evaluation, variation, and genetic control of seed coat whiteness in dry beans (*Phaseolus vulgaris* L.). *J. Am. Soc. Hortic. Sci.* 166, 575–579.
- Kour, A., Boone, A. M., and Vodkin, L. O. (2014). RNA-Seq profiling of a defective seed coat mutation in *Glycine max* reveals differential expression of proline-rich and other cell wall protein transcripts. *PLoS ONE* 9:e96342. doi: 10.1371/journal.pone.0096342
- Krygier, K., Frank Sosulski, F., and Hogge, L. (1982). Free, esterified, and insoluble-bound phenolic acids. 1. Extraction and purification procedure. *J. Agric. Food Chem.* 30, 330–334. doi: 10.1021/jf00110a028
- Kučera, L., Kurka, O., Barták, P., and Bednář, P. (2017). Liquid chromatography/high resolution tandem mass spectrometry - tool for the study of polyphenol profile changes during micro-scale biogas digestion of grape marcs. *Chemosphere* 166, 463–472. doi: 10.1016/j.chemosphere.2016.09.124
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5:11166. doi: 10.1038/srep11166
- Ladizinsky, G. (1998). *Plant Evolution under Domestication*. Dordrecht: Kluwer Academic Publishers.
- Lamport, D. T. A., Kieliszewski, M. J., Chen, Y., and Cannon, M. C. (2011). Role of the Extensin superfamily in primary cell wall architecture. *Plant Physiol.* 156, 11–19. doi: 10.1104/pp.110.169011
- Legesse, N., and Powell, A. A. (1996). Relationship between the development of seed coat pigmentation, seed coat adherence to the cotyledons and the rate of imbibition during the maturation of grain legumes. *Seed Sci. Technol.* 24, 23–32.
- Lenser, T., and Theißen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* 18, 704–714. doi: 10.1016/j.tplants.2013.08.007
- Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252
- Li, P., Chen, B., Zhang, G., Chen, L., Dong, Q., Wen, J., et al. (2016b). Regulation of anthocyanin and proanthocyanidin biosynthesis by *Medicago truncatula* bHLH transcription factor MtTT8. *New Phytol.* 210, 905–921. doi: 10.1111/nph.13816
- Li, P., Dong, Q., Ge, S., He, X., Verdier, J., Li, D., et al. (2016a). Metabolic engineering of proanthocyanidin production by repressing the isoflavone pathways and redirecting anthocyanidin precursor flux in legume. *Plant Biotechnol. J.* 14, 1604–1618. doi: 10.1111/pbi.12524
- Li, Y.-G., Tanner, G., and Larkin, P. (1996). The DMACA-HCl protocol and the threshold proanthocyanidin content for bloat safety in forage legumes. *J. Science Food Agric.* 70, 89–101. doi: 10.1002/(SICI)1097-0010(199601)70:1<89::AID-JSFA470>3.0.CO;2-N
- Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089
- Liljegren, S. J., Roeder, A. H., Kempin, S. A., Gremski, K., Østergaard, L., Guimil, S., et al. (2004). Control of fruit patterning in *Arabidopsis* by INDEHISCENT. *Cell* 116, 843–853. doi: 10.1016/S0092-8674(04)00217-X
- Liu, C., Jun, J. H., and Dixon, R. A. (2014). MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiol.* 165, 1424–1439. doi: 10.1104/pp.114.241877
- Liu, N., Zhang, G., Xu, S., Mao, W., Hu, Q., and Gong, Y. (2015). Comparative transcriptomic analyses of vegetable and grain pea (*Pisum sativum* L.) seed development. *Front. Plant Sci.* 6:1039. doi: 10.3389/fpls.2015.01039
- Lopez-Amoros, M. L., Hernandez, T., and Estrella, I. (2006). Effect of germination on legume phenolic compounds and their antioxidant activity. *J. Food Compos. Anal.* 19, 277–283. doi: 10.1016/j.jfca.2004.06.012
- Ma, F., Cholewa, E., Mohamed, T., Peterson, C. A., and Gijzen, M. (2004). Cracks in the palisade cuticle of soybean seed coats correlate with their permeability to water. *Ann. Bot.* 94, 213–228. doi: 10.1093/aob/mch133
- Marbach, I., and Mayer, A. M. (1974). Permeability of seed coats to water as related to drying conditions and metabolism of phenolics. *Plant Physiol.* 54, 817–820. doi: 10.1104/pp.54.6.817
- Marles, M. A. S., Vandenberg, A., and Bett, K. E. (2008). Polyphenol oxidase activity and differential accumulation of polyphenolics in seed coats of pinto bean (*Phaseolus vulgaris* L.) characterize postharvest color changes. *J. Agric. Food Chem.* 56, 7049–7056. doi: 10.1021/jf8004367
- Matus, A., and Slinkard, A. E. (1993). Effect of fungicidal seed treatment on zero-tannin lentil. *Lens Newsl.* 20, 46–50.
- McCarty, D. R. (1986). A simple method for extraction of DNA from maize tissue. *Maize Genet. Coop. Newsl.* 60, 61.
- McDonald, M. B., Vertucci, C. W., and Roos, E. E. (1988). Soybean seed imbibition: water absorption by seed parts. *Crop Sci.* 28, 993–997. doi: 10.2135/cropsci1988.0011183X002800060026x
- Meyer, C. J., Steudle, E., and Peterson, C. A. (2007). Patterns and kinetics of water uptake by soybean seeds. *J. Exp. Bot.* 58, 717–732. doi: 10.1093/jxb/erl244
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605

- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Miao, Z. H., Fortune, J. A., and Gallagher, J. (2001). Anatomical structure and nutritive value of lupin seed coats. *Aust. J. Agric. Res.* 52, 985–993. doi: 10.1071/AR00117
- Michelmore, R. W., Paran, I., and Kesseli, R. V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl. Acad. Sci. U.S.A.* 88, 9828–9832. doi: 10.1073/pnas.88.21.9828
- Mitsuda, N., and Ohme-Takagi, M. (2008). NAC transcription factors NST1 and NST3 regulate pod shattering in a partially redundant manner by promoting secondary wall formation after the establishment of tissue identity. *Plant J.* 56, 768–778. doi: 10.1111/j.1365-313X.2008.03633.x
- Moïse, J. A., Han, S., Gudynaite-Savitch, L., Johnson, D. A., and Miki, B. L. A. (2005). Seed coats: structure, development, composition, and biotechnology. *In vitro Cell. Dev. Biol. Plant* 41, 620–644. doi: 10.1079/IVP2005686
- Moreau, C., Mike, J., Ambrose, M. J., Turner, L., Hill, L., Ellis, T. H. N., et al. (2012). The b gene of pea encodes a defective flavonoid 3',5'-hydroxylase, and confers pink flower color. *Plant Physiol.* 159, 759–768. doi: 10.1104/pp.112.197517
- Mullin, W. J., and Xu, W. (2000). A study of the intervarietal differences of cotyledon and seed coat carbohydrates in soybean. *Food Res. Int.* 33, 883–891. doi: 10.1016/S0963-9969(00)00118-6
- Mullin, W. J., and Xu, W. (2001). Study of soybean seed coat components and their relationship to water absorption. *J. Agric. Food Chem.* 49, 5331–5335. doi: 10.1021/jf010303s
- Nikolaeva, M. G. (1969). *Physiology of Deep Dormancy in Seeds. Leningrad, Russia, Izdatel'stvo 'Nauka'*. Transl. by Z. Shapiroin Russian. Washington, DC: National Science Foundation.
- Nonogaki, H. (2014). Seed dormancy and germination – emerging mechanisms and new hypotheses. *Front. Plant Sci.* 5:233. doi: 10.3389/fpls.2014.00233
- North, N., Casey, R., and Domoney, C. (1989). Inheritance and mapping of seed lipoxygenase polypeptides in *Pisum*. *Theor. Appl. Genet.* 77, 805–808. doi: 10.1007/BF00268330
- Olsen, K. M., and Wendel, J. F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. *Annu. Rev. Plant Biol.* 64, 47–70. doi: 10.1146/annurev-arplant-050312-120048
- Patil, G., Valliyodan, B., Deshmukh, R., Prince, S., Nicander, B., Zhao, M., et al. (2015). Soybean (*Glycin max*) SWEET gene family: insight through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics* 16:520. doi: 10.1186/s12864-015-1730-y
- Pazhamala, L. T., Agarwal, G., Bajaj, P., Kumar, V., Kulshreshtha, A., Saxena, R. K., et al. (2016). Deciphering transcriptional programming during pod and seed development using RNA-Seq in pigeonpea (*Cajanus cajan*). *PLoS ONE* 11:e0164959. doi: 10.1371/journal.pone.0164959
- Pizzi, A., and Cameron, F. A. (1986). Flavonoid tannins — structural wood components for drought-resistance mechanisms of plants. *Wood Sci. Technol.* 20:119.
- Porter, L. J. (1989). “Condensed tannins,” in *Natural Products of Woody Plants I: Chemicals Extraneous to the Lignocellulosic Cell*, ed J. W. Rowe (Berlin: Springer), 651–690. doi: 10.1007/978-3-642-74075-6_18
- Pradhan, S., Bandhiwal, N., Shah, N., Kant, C., Gaur, R., and Bhatia, S. (2014). Global transcriptome analysis of developing chickpea (*Cicer arietinum* L.) seeds. *Front. Plant Sci.* 5:698. doi: 10.3389/fpls.2014.00698
- Purugganan, M. D., and Fuller, D. Q. (2009). The nature of selection during plant domestication. *Nature* 457, 843–848. doi: 10.1038/nature07895
- Radchuk, V., and Borisjuk, L. (2014). Physical, metabolic and developmental functions of the seed coat. *Front. Plant Sci.* 5:510. doi: 10.3389/fpls.2014.00510
- Ramsay, G. (1997). Inheritance and linkage of a gene for testa imposed seed dormancy in faba bean (*Vicia faba* L.). *Plant Breed.* 116, 287–289. doi: 10.1111/j.1439-0523.1997.tb00998.x
- Ranal, M. A., and Santana, D. G. (2006). How and why to measure the germination process? *Rev. Bras. Bot.* 29, 1–11. doi: 10.1590/S0100-84042006000100002
- Ranathunge, K., Shao, S., Qutob, D., Gijzen, M., Peterson, C. A., and Bernards, M. A. (2010). Properties of the soybean seed coat cuticle change during development. *Planta* 231, 1171–1188. doi: 10.1007/s00425-010-1118-9
- Redekar, N. R., Biyashev, R. M., Jensen, R. V., Helm, R. F., Grabau, E. A., and Maroof, M. A. (2015). Genome-wide transcriptome analyses of developing seeds from low and normal phytic acid soybean lines. *BMC Genomics* 16:1074. doi: 10.1186/s12864-015-2283-9
- Righetti, K., Vu, J. L., Pelletier, S., Vu, B. L., Glaab, E., Lalanne, D., et al. (2015). Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. *Plant Cell* 27, 2692–2708. doi: 10.1105/tpc.15.00632
- Robert, C., and Watson, M. (2015). Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biol.* 16:177. doi: 10.1186/s13059-015-0734-x
- Roth, A., Ding, J., Morin, R., Crisan, A., Ha, G., Giuliany, R., et al. (2012). JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* 28, 907–913. doi: 10.1093/bioinformatics/bts053
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Seiki, M. (1999). Membrane-type matrix metalloproteinases. *Acta Pathol. Microbiol. Immunol. Scand.* 107, 137–143.
- Severin, A., Woody, J. L., Bolon, Y.-T., Joseph, B., Diers, B. W., Farmer, A. D., et al. (2010). RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.* 10:160. doi: 10.1186/1471-2229-10-160
- Shao, S., Meyer, C. J., Ma, F., Peterson, C. A., and Bernards, M. A. (2007). The outermost cuticle of soybean seeds: chemical composition and function during imbibition. *J. Exp. Bot.* 58, 1071–1082. doi: 10.1093/jxb/erl268
- Shi, J., and Lai, J. (2015). Patterns of genomic changes with crop domestication and breeding. *Curr. Opin. Plant Biol.* 24, 47–53. doi: 10.1016/j.pbi.2015.01.008
- Singh, R., Singh, S., Parihar, P., Mishra, R. K., Tripathi, D. K., Singh, V. P., et al. (2016). Reactive oxygen species (ROS): beneficial companions of plant's developmental processes. *Front. Plant Sci.* 7:1299. doi: 10.3389/fpls.2016.01299
- Smartt, J. (1990). *Grain Legumes: Evolution and Genetic Resources*. Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511525483
- Smykal, P., Aubert, G., Burstin, J., Coyne, C. J., Ellis, N. T. H., Flavell, A. J., et al. (2012). Pea (*Pisum sativum* L.) in the genomic era. *Agronomy* 2, 74–115. doi: 10.3390/agronomy2020074
- Smykal, P., Coyne, C. J., Ambrose, M. J., Maxted, N., Schaefer, H., Blair, M. W., et al. (2015). Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev. Plant Sci.* 34, 43–104. doi: 10.1080/07352689.2014.897904
- Smykal, P., Vernoud, V., Blair, M. W., Soukup, A., and Thompson, R. D. (2014). The role of the testa during development and in establishment of dormancy of the legume seed. *Front. Plant Sci.* 5:351. doi: 10.3389/fpls.2014.00351
- Soukup, A. (2014). “Selected simple methods of plant cell wall histochemistry and staining for light microscopy,” in *Plant Cell Morphogenesis: Methods and Protocols, Vol. 1080, Methods in Molecular Biology*, eds V. Žársky and F. Cvrčková (New York, NY: Springer), 25–40. doi: 10.1007/978-1-62703-643-6_2
- Soukup, A., and Tylová, E. (2014). “Essential methods of plant sample preparation for light microscopy,” in *Plant Cell Morphogenesis: Methods and Protocols, Vol. 1080, Methods in Molecular Biology*, eds V. Žársky and F. Cvrčková (New York, NY: Springer), 1–23. doi: 10.1007/978-1-62703-643-6_1
- Sreerama, N. Y., Sashikala, V. B., and Pratapa, V. M. (2010). Variability in the distribution of phenolic compounds in milled fractions of chickpea and horse gram: evaluation of their antioxidant properties. *J. Agric. Food Chem.* 58, 8322–8330. doi: 10.1021/jf101335r
- Suanum, W., Somta, P., Kongjaimun, A., Yimram, T., Kage, A., Tomooka, N., et al. (2016). Co-localization of QTLs for pod fiber content and pod shattering in F-2 and backcross populations between yardlong bean and wild cowpea. *Mol. Breeding* 36, 80. doi: 10.1007/s11032-016-0505-8
- Sudheesh, J., Sawbridge, T. I., Cogan, N. O., Kennedy, P., Forster, J. W., and Kaur, S. (2015). *De novo* assembly and characterization of field pea transcriptome using RNA-Seq. *BMC Genomics* 16:611. doi: 10.1186/s12864-015-1815-7
- Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y., et al. (2015). GmHs1-1, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nat. Genet.* 47, 939–943. doi: 10.1038/ng.3339
- Suzuki, M., Fujino, K., Nakamoto, Y., Ishimoto, M., and Funatsuki, H. (2010). Fine mapping and development of DNA markers for the *qPDH1* locus associated with pod dehiscence in soybean. *Mol. Breed.* 25, 407–418. doi: 10.1007/s11032-009-9340-5
- Tayeh, N., Aluome, C., Falque, M., Jacquin, F., Klein, A., Chauveau, A., et al. (2015). Development of two major resources for pea genomics: the GenoPea 13.2K SNP

- Array and a high density, high resolution consensus genetic map. *Plant J.* 84, 1257–1273. doi: 10.1111/tpl.13070
- Tiwari, S. P., and Bhatia, V. S. (1995). Characters of pod anatomy associated with resistance to pod-shattering in soybean. *Ann. Bot.* 76, 483–485. doi: 10.1006/anbo.1995.1123
- Troszyńska, A., and Ciska, E. (2002). Phenolic compounds of seed coats of white and coloured varieties of pea (*Pisum sativum* L.) and their total antioxidant activity. *Czech J. Food Sci.* 20, 15–22.
- Trygg, J., and Wold, S. (2002). Orthogonal projections to latent structures (O-PLS). *J. Chemometrics* 16, 119–128. doi: 10.1002/cem.695
- Tuteja, J. H., Clough, S. J., Chan, W. C., and Vodkin, L. O. (2004). Tissue specific gene silencing mediated by naturally occurring chalcone synthase gene cluster in *Glycine max*. *Plant Cell* 16, 819–835. doi: 10.1105/tpc.021352
- Verdier, J., Dessaint, F., Schneider, C., and Abirached-Darmency, M. (2013a). A combined histology and transcriptome analysis unravels novel questions on *Medicago truncatula* seed coat. *J. Exp. Bot.* 64, 459–470. doi: 10.1093/jxb/ers304
- Verdier, J., Torres-Jerez, I., Wang, M., Andriankaja, A., Allen, S. N., He, J., et al. (2013b). Establishment of the *Lotus japonicus* gene expression atlas (LjGEA) and its use to explore legume seed maturation. *Plant J.* 74, 351–362. doi: 10.1111/tpl.12119
- Vu, D. T., Velusamy, V., and Park, E. (2014). Structure and chemical composition of wild soybean seed coat related to its permeability. *Pak. J. Bot.* 46, 1847–1857.
- Wan, L., Li, B., Pandey, M. K., Wu, Y., Lei, Y., Yan, L., et al. (2016). Transcriptome analysis of a new peanut seed coat mutant for the physiological regulatory mechanism involved in seed coat cracking and pigmentation. *Front. Plant Sci.* 7:1491. doi: 10.3389/fpls.2016.01491
- Wang, C. S., and Vodkin, L. O. (1994). Extraction of RNA from tissues containing high levels of proanthocyanidins that bind RNA. *Plant Mol. Biol. Rep.* 12, 132–145.
- Wang, H. L., and Grusak, M. A. (2005). Structure and development of *Medicago truncatula* pod wall and seed coat. *Ann. Bot.* 95, 737–747. doi: 10.1093/aob/mci080
- Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics* 26, 136–138. doi: 10.1093/bioinformatics/btp612
- Weeden, N. F. (2007). Genetic changes accompanying the domestication of *Pisum sativum*: is there a common genetic basis to the Domestication syndrome for legumes? *Ann. Bot.* 100, 1017–1025. doi: 10.1093/aob/mcm122
- Weeden, N. F., Brauner, S., and Przyborowski, J. A. (2002). Genetic analysis of pod dehiscence in pea (*Pisum sativum* L.). *Cell. Mol. Biol. Lett.* 7, 657–663.
- Weitbrecht, K., Muller, K., and Leubner-Metzger, G. (2011). First off the mark: early seed germination. *J. Exp. Bot.* 62, 3289–3309. doi: 10.1093/jxb/err030
- Werker, E., Marbach, I., and Mayer, A. M. (1979). Relation between the anatomy of the testa, water permeability and the presence of phenolics in the genus *Pisum*. *Ann. Bot.* 43, 765–771. doi: 10.1093/oxfordjournals.aob.a085691
- Xie, D. Y., Sharma, S. B., and Dixon, R. A. (2004). Anthocyanidin reductases from *Medicago truncatula* and *Arabidopsis thaliana*. *Arch. Biochem. Biophys.* 422, 91–102. doi: 10.1016/j.abb.2003.12.011
- Xie, D. Y., Sharma, S. B., Paiva, N. L., Ferreira, D., and Dixon, R. A. (2003). Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. *Science* 299, 396–399. doi: 10.1126/science.1078540
- Yang, K., Jeong, N., Moon, J., Lee, Y. H., Lee, S. H., Kim, H. M., et al. (2010). Genetic analysis of genes controlling natural variation of seed coat and flower colors in soybean. *J. Heredity* 101, 757–768. doi: 10.1093/jhered/esq078
- Yin, Q., Shen, G., Chang, Z., Tang, Y., Gao, H., and Pang, Y. (2017). Involvement of three putative glucosyltransferases from the UGT72 family in flavonol glucoside/rhamnoside biosynthesis in *Lotus japonicus* seeds. *J. Exp. Bot.* 68, 597–612. doi: 10.1093/jxb/erw420
- Young, M. D., Wakefield, M. J., Smyth, G. K., and Oshlack, A. (2010). Gene ontology analysis for RNA-Seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14
- Zawada, A. M., Rogacev, K., Müller, S., Rotter, B., Winter, P., Fliser, D., et al. (2014). Massive analysis of cDNA ends (MACE) and miRNA expression profiling identifies proatherogenic pathways in chronic kidney disease. *Epigenetics* 9, 161–172. doi: 10.4161/epi.26931
- Zhang, S., Shi, Y., Cheng, N., Du, H., Fan, W., and Wang, C. (2015). *De novo* characterization of fall dormant and nondormant alfalfa (*Medicago sativa* L.) leaf transcriptome and identification of candidate genes related to fall dormancy. *PLoS ONE* 10:e0122170. doi: 10.1371/journal.pone.0122170
- Zhao, J., Pang, Y., and Dixon, R. A. (2010). The mysteries of proanthocyanidin transport and polymerization. *Plant Physiol.* 153, 437–443. doi: 10.1104/pp.110.155432
- Zhao, Q., and Dixon, R. A. (2011). Transcriptional networks for lignin biosynthesis: more complex than we thought? *Trends Plant Sci.* 16, 227–233. doi: 10.1016/j.tplants.2010.12.005
- Zhao, S., Tuan, P. A., Li, X., Kim, Y. B., Kim, H., Park, C. G., et al. (2013). Identification of phenylpropanoid biosynthetic genes and phenylpropanoid accumulation by transcriptome analysis of *Lycium chinense*. *BMC Genomics* 14:802. doi: 10.1186/1471-2164-14-802
- Zhong, R., Lee, C., and Ye, Z. (2010). Evolutionary conservation of the transcriptional network regulating secondary cell wall biosynthesis. *Trend Plant Sci.* 15, 625–632. doi: 10.1016/j.tplants.2010.08.007
- Zhou, S., Sekizuka, H., Yang, Z., Sawa, S., and Pan, J. (2010). Phenolics in the seed coat of wild soybean (*Glycine soja*) and their significance for seed hardness and seed germination. *J. Agric. Food Chem.* 58, 10972–10978. doi: 10.1021/jf102694k
- Zhou, Z., Jiang, Y., Wang, Z., Gou, Z., Lyu, J., Li, W., et al. (2015). Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nature Biotech.* 33, 408–414. doi: 10.1038/nbt.3096
- Zhu, W., Chen, X., Li, H., Zhu, F., Hong, Y., Varshney, R. K., et al. (2014). Comparative transcriptome analysis of aerial and subterranean pods development provides insights into seed abortion in peanut. *Plant Mol. Biol.* 85, 395–409. doi: 10.1007/s11103-014-0193-x
- Zhukov, V. A., Zhernakov, A. I., Kulaeva, O. A., Ershov, N. I., Borisov, A. Y., and Tikhonovich, I. A. (2015). *De novo* assembly of the Pea (*Pisum sativum* L.) nodule transcriptome. *Int. J. Genomics* 2015:695947. doi: 10.1155/2015/695947
- Zohary, D., and Hopf, M. (2000). *Domestication of Plants in the Old World, 3rd Edn.* New York, NY: Oxford University Press.
- Zou, H., Tzarfati, R., Hübner, S., Krugman, T., Fahima, T., Abbo, S., et al. (2015). Transcriptome profiling of wheat glumes in wild emmer, hulled landraces and modern cultivars. *BMC Genomics* 16:777. doi: 10.1186/s12864-015-1996-0

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Hradilová, Trněný, Váľková, Cechová, Janská, Prokešová, Aamir, Krezdorn, Rotter, Winter, Varshney, Soukup, Bednář, Hanáček and Smýkal. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Major Contribution of Flowering Time and Vegetative Growth to Plant Production in Common Bean As Deduced from a Comparative Genetic Mapping

Ana M. González¹, Fernando J. Yuste-Lisbona², Soledad Saburido³, Sandra Bretones², Antonio M. De Ron¹, Rafael Lozano² and Marta Santalla^{1*}

¹ Grupo de Biología de Agrosistemas, Misión Biológica de Galicia-Consejo Superior de Investigaciones Científicas, Pontevedra, Spain, ² Departamento de Biología y Geología (Genética), Centro de Investigación en Biotecnología Agroalimentaria, Universidad de Almería, Almería, Spain, ³ Somos Semilla, Seed Library, Guanajuato, Mexico

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Spanish National Research Council
(CSIC), Spain

Reviewed by:

Elisa Bellucci,
Marche Polytechnic University, Italy
Moustafa Bani,
University of Córdoba, Spain

*Correspondence:

Marta Santalla
msantalla@mbg.csic.es

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 29 July 2016

Accepted: 07 December 2016

Published: 26 December 2016

Citation:

González AM, Yuste-Lisbona FJ, Saburido S, Bretones S, De Ron AM, Lozano R and Santalla M (2016) Major Contribution of Flowering Time and Vegetative Growth to Plant Production in Common Bean As Deduced from a Comparative Genetic Mapping. *Front. Plant Sci.* 7:1940. doi: 10.3389/fpls.2016.01940

Determinacy growth habit and accelerated flowering traits were selected during or after domestication in common bean. Both processes affect several presumed adaptive traits such as the rate of plant production. There is a close association between flowering initiation and vegetative growth; however, interactions among these two crucial developmental processes and their genetic bases remain unexplored. In this study, with the aim to establish the genetic relationships between these complex processes, a multi-environment quantitative trait locus (QTL) mapping approach was performed in two recombinant inbred line populations derived from inter-gene pool crosses between determinate and indeterminate genotypes. Additive and epistatic QTLs were found to regulate flowering time, vegetative growth, and rate of plant production. Moreover, the pleiotropic patterns of the identified QTLs evidenced that regions controlling time to flowering traits, directly or indirectly, are also involved in the regulation of plant production traits. Further QTL analysis highlighted one QTL, on the lower arm of the linkage group Pv01, harboring the *Phvul.001G189200* gene, homologous to the *Arabidopsis thaliana* *TERMINAL FLOWER1* (*TFL1*) gene, which explained up to 32% of phenotypic variation for time to flowering, 66% for vegetative growth, and 19% for rate of plant production. This finding was consistent with previous results, which have also suggested *Phvul.001G189200* (*PvTFL1y*) as a candidate gene for determinacy locus. The information here reported can also be applied in breeding programs seeking to optimize key agronomic traits, such as time to flowering, plant height and an improved reproductive biomass, pods, and seed size, as well as yield.

Keywords: *Phaseolus vulgaris* L., flowering time, vegetative growth, rate of plant production, quantitative trait locus

INTRODUCTION

Common bean (*Phaseolus vulgaris* L.) is the most important food legume for direct human consumption. It is considered to be a rich source of proteins, micronutrients, and calories for human daily needs (Broughton et al., 2003). It is grown over a wide range of latitudes; however, there is an adaptation of each cultivar to a relatively narrow range of latitudes. Wild bean forms present indeterminate growth habits and require day-lengths of <12 h to initiate flowering. Together, earlier flowering and determinate growth habit genotypes under day-lengths longer than 12 h allowed for the adaptation to higher latitudes (Gepts and Debouck, 1991). Domesticated and wild common bean display notable differences in growth habit types. Thus, there is a considerable variability in domesticated cultivars, which show pronounced differences in growth form, i.e., determinate vs. indeterminate, and growth habit, i.e., I, II, III, and IV (Debouck and Hidalgo, 1986; Debouck, 1991). Determinate common bean cultivars generally flower and mature early, and the transition of the terminal shoot meristem from vegetative to reproductive state results in a terminal inflorescence in the axil of the older leaf primordia. By contrast, in indeterminate cultivars, the terminal shoot meristem continuously produces modular units until senescence, each one consisting of a leaf and an inflorescence. Thus, the plant will have a terminal shoot meristem that remains in a vegetative state throughout the production of vegetative and reproductive structures (Ojehomon and Morgan, 1969; Tanaka and Fujita, 1979). It has been documented that stem termination has great effects on plant height, flowering and maturity period, amount of branching, length of internodes on the main stem, and node production, which conditions how many flowers and leaves, and therefore pods and seeds, are produced. Thus, understanding the genetic control of vegetative growth and flowering time in common bean will enable genetic manipulation of major components of yield.

Previous studies demonstrated that the *FIN* locus mainly regulates stem growth habit in common bean and that the indeterminate growth habit is dominant over the determinate one. This gene was mapped on the Linkage Group (LG) Pv01 (Norton, 1915; Koinange et al., 1996). Despite the fact that *FIN* is a monogenic locus, it is possible to find a wide range of stem termination types among common bean cultivars, which may be regulated by a second unnamed locus mapped on Pv07 (Kolkman and Kelly, 2003). In addition, control of twining has been attributed to the *TOR* gene, distinct from the *FIN* locus (Norton, 1915) although either *FIN* has a pleiotropic effect on twining or that *TOR* is tightly linked to *FIN*. Additionally, other loci have been reported to control flowering time and other flowering-related traits in common bean (Norton, 1915; Wallace et al., 1993; Jung et al., 1996; Bassett, 1997; McClean et al., 2002; Tar'an et al., 2002; Kolkman and Kelly, 2003; Blair et al., 2006; Checa and Blair, 2008; Chavarro and Blair, 2010; Pérez-Vega et al., 2010). The majority of these loci have been detected as Quantitative Trait Loci (QTL), although they have been treated as Mendelian factors in some studies. Probably, different works have detected the same loci; however,

the lack of common markers among different mapping studies makes difficult to determine whether or not they are the same locus.

During the past two decades, the model species *Arabidopsis thaliana* has been mainly used to study the process of phase transition from vegetative to reproductive growth at developmental, environmental, genetic, and molecular levels (Weigel, 1995; Yanofsky, 1995; Bradley et al., 1997; Ma, 1998; Pidkowich et al., 1999). In this species, floral meristem identity is determined by two different pathways. The heterodimer FLOWERING LOCUS T (FT)/FLOWERING LOCUS D (FD) is proposed to promote flowering initiation through activating the *APETALA1* (*API*) expression (Abe et al., 2005; Wigge et al., 2005). Furthermore, another key regulator of flowering initiation is *TERMINAL FLOWER1* (*TFL1*), a floral repressor and a regulator of inflorescence meristem development, which acts by repressing the expression of *API* and *LEAFY* (*LFY*) floral identity genes (Bradley et al., 1997; Ohshima et al., 1997; Nilsson et al., 1998; Boss et al., 2004). FT and TFL1 have closely related sequences, although key amino acids have been found, which have been proposed as responsible for making that these two proteins perform opposite functions. Within the legumes, the value of a comparative approach to candidate gene identification has led to the characterization of the molecular identity of *TFL1* co-orthologs such as *FIN* in common bean (designated as *PvTFL1y*; Repinski et al., 2012); *Dt1* in soybean (*GmTFL1*; Tian et al., 2010); and *DETERMINATE* (*DET*; *PsTFL1a*) and *LATE FLOWERING* (*LF*; *PsTFL1c*) in pea (Foucher et al., 2003; Weller and Ortega, 2015). Recent findings have clearly shown that in several legume species, determinate inflorescence architecture is conferred by mutation of specific *TFL1* genes (Benlloch et al., 2015). The determinate growth habit caused by mutations within specific *TFL1*-homologs in other grain legumes indicates that the determinate function is conserved in these species (Kong et al., 2010; Tian et al., 2010; Kwak et al., 2012; Repinski et al., 2012; Dhanasekar and Reddy, 2014).

In this study, the aim was to identify the genetic determinants of vegetative growth and its relationship to days to flowering and fruit maturation in two Recombinant Inbred Line (RIL) populations by using a mixed-model based composite mapping method for QTL detection. Both populations were derived from inter-gene (Andean × Mesoamerican) pool crosses between determinate and indeterminate genotypes, and shared an Andean determinate common parent, which allowed for the comparison of the results in reference to a tester line. Parents of each RIL population also differ in rate of plant production (leaf, flower and fruit size, and yield components), allowing for the dissection of the genetic architecture of these traits, as well as the study of the relationship among vegetative growth and these traits. Comparative QTL mapping indicated that, in both RIL mapping populations, the genomic region where the *FIN* locus is located not only is involved in the regulation of vegetative growth and rate of plant production but also affect flowering time, suggesting a pattern of pleiotropic effects accounting for the genetic bases of these traits.

MATERIALS AND METHODS

Plant Material and Trials

Two F_{2:8} RIL populations derived by single-seed descent from an F₂ population from crosses between the Mesoamerican (M) and Andean (A) gene pools were used. The MA population was generated from the cross between lines PHA0419 (Mesoamerican) and Beluga (Andean), whereas the AM population was obtained from the Beluga and PHA0399 (Mesoamerican) cross. Beluga is a large white kidney seed which seed weight averages 60 g 100 seeds⁻¹ (range: 52–66 g 100 seeds⁻¹), 55 cm in length of main stem with a type I determinate growth habit, and blooms 30 days after planting (DAP). Beluga is resistant to bean common mosaic virus (BCMV), as it bears the autosomal dominant hypersensitive *I* gene, and possesses the Andean *Co-1* gene for resistance to races 65 and 73 of anthracnose (Kelly et al., 1999). Both PHA0419 and PHA0399 are great northern beans which averages 86 g 100 seeds⁻¹ (range: 76–103 g 100 seeds⁻¹), and shared a type IV climbing indeterminate growth habit (averages 278 cm of length of main stem) and late flowering (46 DAP as average). PHA-0419 possesses the Mesoamerican *Co-4²*, *Co-6*, and *Co-10* genes that condition resistance to races 23, 39, 102, 448, and 1545 of anthracnose, while PHA-0399 carries *Co-4²*, *Co-4³*, and *Co-6* genes that condition resistance to races 23, 55, 102, and 1545 of anthracnose (Santalla et al., 2004; **Figure 1**). RILs and parents were evaluated in open field (F) and greenhouse (G) environments in Pontevedra (Spain, 42°24' N, 8°38' W, 40 masl). Briefly, the RILs populations and the parents were planted between April and May in 2008 and 2009 under field conditions: F108 planted on 2nd April 2008 (93 Julian days and day-length ~12.4 h) and F109 planted on 8th May 2009 (128 Julian days and day-length ~11.1 h). RIL and parental lines were also planted between February and March in 2009 and 2008 under greenhouse conditions: G108 planted on 25th March 2008 (85 Julian days and day-length ~12.2 h) and G109 planted on 9th February 2009 (40 Julian days and day-length ~10.1 h). Field and greenhouse experiments were designed according to a complete randomized block design with two replications. Each genotype was planted in a single row plot, 0.80 m apart and 3.0 m long with a total of 15 plants per row; the density was 30,000 plants/ha. Crop management was in accordance with common bean local practices.

Field Evaluation and Data Collection

Data were collected for time to flowering and fruit maturity evaluated as days to flowering, days to immature pod harvest, and days to physiological pod maturity (Supplementary Table 1). The days to flowering (FT) trait was scored as the number of days from planting date to the opening of the first flower. The days to young-green pod (PGT) trait was recorded when a plant had 50% of the immature pods. The days to physiological pod maturity (PST) trait was scored as the number of days from planting date to the appearance of a mature dry pod on a primary branch.

Measurements were collected for growth ability reordered as length of main stem, number of primary stem branches, and internode length, in five random plants from each RIL. Length of main stem (LMS) was measured in cm from the base of the

plant to the uppermost leaflet on the longest branch. The number of primary stem branches (NPB) was recorded as the number of the stems winding around the support strings at the midpoint of the plant length. Internode length (LI) was recorded at the fifth internode on the main stem in cm.

To quantify crop growth and productivity, flower, pod and seed size, number of seeds per pod, number of pods per plant, and seed yield were determined in ten random flowers and fruits on a plot basis. Bracteole length (BL) measured in mm along the midrib of the lamina. Bracteole width (BWI) recorded in mm between the widest lobes of the lamina perpendicular to the lamina mid-rib. Maximum leaflet width (LWI) was measured in cm at the largest point perpendicular to the midrib. Leaflet length (LL) was recorded in mm from the lamina tip to the point of petiole intersection along the midrib. Pod length (PL) was recorded in mm as the exterior distance from the peduncle connection point to the apex excluding the beak. Pod width (PWI) determined in mm as the distance at right angles to the sutures at the level of the second seed from the apex. Pod thickness (PT) was recorded in mm as the distance between sides at the level of the second and third seed from the apex. Seed width (SWI) was the longest distance across the hilum, in mm. Seed thickness (ST) was measured in mm from hilum to opposite side. Seed length (SL) was measured in mm parallel to the hilum. Seed weight (SW) was determined on 100 dry seeds per plot. The number of seeds per pod (NSP), number of pods per plant (NPP), and seed yield (SY) were recorded and expressed in kilograms per hectare at moisture content of 140 g kg⁻¹.

Experimental Design and Statistical Data Analysis

Comprehensive statistical analysis (mean value, standard error, and range of variation) and normality test (Kolmogorov-Smirnov test) were carried out for each quantitative trait and environment. LMS, NPB, LI, BL, BWI, SW, NSP, NPP, and SY traits failed to meet normality assumptions and the Box-Cox transformation was used to improve normality, while LOG transformation was used for LL and LWI traits. Significant variation in the expression of traits through the environment conditions was analyzed using PROC MIXED (SAS Institute Inc. 9.04, Cary, NC, USA). The estimates of variance components were obtained by the REML method with Proc MIXED in SAS9.04 and used to calculate the broad-sense heritability on a progeny-mean basis ($h^2 = \sigma_\lambda^2 / [\sigma_\lambda^2 / e + \sigma_\lambda^2 + (\sigma_e^2 / re)]$ where: σ_λ^2 = genetic variance of the trait; σ_e^2 = variance due to environmental factors; σ_e^2 = error variance; r = number of replications and e = number of environments). In order to increase the precision of the entry mean basis heritability estimate, it was used the harmonic means for the number of replications and environments, where each experimental line was tested (Holland et al., 2003). Approximate standard errors of heritability estimates were obtained by the delta method (Holland, 2006). Phenotypic Pearson correlation coefficients for the different traits were calculated using PROC CORR through the environment conditions in SAS9.04.



FIGURE 1 | Determinate type I and indeterminate type IV growth habits, and dry seeds of the parental genotypes Beluga (A,C), PHA0399 (B,D), and PHA0419 (B,E). Scale bar 1 cm.

DNA Isolation and Marker Analysis

Young leaves from individual plants of both RIL mapping populations (60 and 179 lines for AM and MA, respectively) were collected for genomic DNA isolation as described by Chen and Ronald (1999). Total DNA was stored in sterile water and examined by electrophoresis in 1% agarose gels in 1X SB buffer (10 mM sodium boric acid). The concentration and quality of extracted DNA was determined by reading at 230, 260, and 280 nm using Nanodrop Thermo Scientific™ (NanoDrop 2000). DNA was diluted in sterile water to get a working dilution of 5–10 ng/μL, which was used for PCR analysis.

A set of 634 markers [Simple Sequence Repeats (SSR) and single nucleotide polymorphisms (SNP)] were tested for polymorphisms in the parental genotypes. SSR markers were named according to the respective authors [IAC: Benchimol et al., 2007; Cardoso et al., 2008; Campos et al., 2011; BM, GATS: Gaitán-Solís et al., 2002; Blair et al., 2009a; BMb: Córdoba

et al., 2010; BMc: Blair et al., 2009b, 2011; BMd: Blair et al., 2003; PV, Pvtttc001: Yu et al., 2000; PvBR: Buso et al., 2006; Grisi et al., 2007; PVEST, X04001: García et al., 2011; PvM: Hanai et al., 2007]. SSR markers were evaluated either by gel electrophoresis or capillary electrophoresis in an ABI PRISM® 3130 XL Genetic Analyzer (Applied Biosystems, USA). SNP markers were analyzed by High Resolution Melting Technology (HRM) using a LightScanner instrument (Idaho Technology), according to the protocols described by Montgomery et al. (2007). These markers were designated as Leg- (Hougaard et al., 2008) and SNP- (McConnell et al., 2010), respectively.

Linkage Map and QTL Analysis

JoinMap 4.0 software (van Ooijen, 2006) was used to construct the genetic linkage maps for both MA and AM mapping populations. A minimum logarithm of odds ratio (LOD) of 6.0 was considered to establish significant linkage. Locus order

within the LOD grouping was generated for each LG using the regression mapping algorithm with the following JoinMap parameters: Rec = 0.3, LOD = 2.0, and Jump = 5. The Kosambi map function (Kosambi, 1944) was used to calculate the genetic distance between markers. LGs were designated according to Pedrosa-Harand et al. (2008). JoinMap 4.0 (van Ooijen, 2006) was also used to generate pairwise recombination frequencies and LOD scores for the selected sets of representative loci for each LG, which were then combined into a single group node in the navigation tree. The regression mapping algorithm was used and the LG lengths for the consensus map of all the representative markers were calculated.

The physical position of genetic markers was obtained by sequence similarity analysis using BLASTN (Altschul et al., 1997) against the common bean genome (Phytozome v.11: Pv1.0; Schmutz et al., 2014) in the Phytozome database (<http://www.phytozome.net/>). The correlations between physical distance and genetic map in each LG were calculated by Spearman's rank correlation coefficients.

QTLNetwork 2.0 software (Yang et al., 2008) was used for multi-environment QTL analyses. In order to identify putative single-locus QTLs and their environment interactions (QTLs \times Environment, QE), a mixed-model based composite interval mapping method (MCIM) was carried out for one-dimensional genome scan. In addition, with the aim to detect epistatic QTLs (E-QTL) and their environment interaction effects (E-QTLs \times Environment, E-QE), a two-dimensional genome scan was performed. A QTL was declared significant when the *F-value* was higher than the *F-value* threshold determined by a 1000 permutation test at 95% confidence level. Markov Chain Monte Carlo method was used to estimate the effects of QTLs and environment interactions (Wang et al., 1994). Both testing and filtration window sizes were set at 10 cM, with a walk speed of 1 cM. Candidate interval selection, putative QTL detection, and QTL effect was estimated with an experimental-wise significance level of 0.05. MapChart 2.2 software (Voorrips, 2002) was used to draw the genetic map and the detected QTLs. QTL regions were positioned onto the consensus map. QTL designations were made using abbreviations for the quantitative trait, and followed by LG number at which the QTL was mapped.

RESULTS

Vegetative Growth and Time to Flowering Variation

Genetic variation for vegetative growth as length of main stem and rate of plant production has been studied, together with its relationship to days to flowering and maturity in both inter-gene pool RIL populations. The populations were grown in different years and locations (field vs. greenhouse). Both populations segregated for different levels of growth ability: indeterminate vs. determinate growth habits. Classification of the 60 and 179 lines of AM and MA RIL populations for growth habit identified 37 and 63 lines as homozygous determinate type I, and 23 and 116 lines as homozygous indeterminate type IV, respectively. The observed growth habit segregation fitted to a

1:1 ratio ($\chi^2 = 3.27$, $P \leq 0.05$) for the AM population, indicating that a single gene determined the trait. However, growth habit distribution appeared distorted in the MA population ($\chi^2 = 15.69$, $P \geq 0.05$). On the basis of segregation analysis results, the gene for growth habit (*FIN*) was mapped along with the molecular markers. Supplementary Figure 1 shows phenotypic distribution of the RIL populations based on line means. The large range of variation and transgressive segregations observed for most traits in both RIL populations suggested a complex control of these traits, with positive alleles shared between the two parents of the RIL populations. Transgressive segregation in both directions was observed for days to flowering and maturity in both populations. For LMS, bimodality was observed in the AM population, with a clear separation of phenotypic classes that would indicate monogenic inheritance, although this separation was not evident in the MA distribution. Lines shorter or taller than the height of the parents were found for LMS trait in both populations. Likewise, for NPB, the number of branches produced by many of the RILs was higher than the parental lines in both populations, which indicated a positive transgressive segregation for this trait, although some skewing was observed in MA to low values. For LI, the histograms showed a similar pattern across both populations, indicating positive transgressive segregation. Hence, in general, the phenotypic segregations for rate of production traits in these two RIL populations exhibited normal distribution, and transgressive segregation in both directions, a typical phenomenon of a quantitative trait, regulated by several genes and influenced by the environment.

Mean values, standard errors, and ranges of variation for the quantitative traits in each RIL mapping population for each environmental condition have been summarized in Supplementary Tables 2, 3. Mesoamerican PHA0419 and PHA0399 lines were late in flowering and taller, with larger rates of plant production compared to Beluga. In both RIL populations and for all evaluated traits, it was found differences among environments in mean values and ranges of variation, although environment \times line interactions were not significant for most of the evaluated traits in both RIL populations. Significant differences among parents were detected except in some environments for PST, LL, LWI, PL, and PWI traits in the MA RIL population (Supplementary Table 2) and for BL, PWI, and ST traits in the AM RIL population (Supplementary Table 3). Despite of that, it was observed significant differences among RILs for all quantitative traits except for LL in MA population under F108 and F109 environmental conditions (Supplementary Table 2).

High broad-sense heritability estimates (≥ 0.50 ; Supplementary Table 4) were detected for most of the quantitative traits across the four environments except for NPB, LI and BWI in both populations, and PST, BL, and PT in MA population. Higher heritability estimates and correlations were observed in AM population than in MA population. LMS and FT were significantly correlated in both AM ($r = 0.50$, $P \leq 0.001$) and MA ($r = 0.38$, $P \leq 0.001$) populations. Their inter-relationship suggests that some genomic regions influence both traits. Finally, a negative and significant correlation was observed between FT and NPB ($r = -0.43$, $P \leq 0.001$) and

positive and significant correlation between FT and LI ($r = 0.33$, $P \leq 0.001$) in AM population, while no significant correlation values were observed in MA population. These correlations indicated that, on average, genotypes with a low number of primary branches, and high length of internodes showed a later flowering date and a longer main stem. As expected from the ontogenetic pattern of vegetative growth, LMS-values were positive and significant correlated with plant parts size (bracteole, leaf, pod, and seed) in both populations except for BWI in AM RIL, suggesting a pleiotropic effect. Positive and significant correlations were observed between FT, PGT, and PST; these last traits were positively correlated with LI in both RIL populations. No significant or low correlation values were observed between FT and all other rate of plant production traits measured. Significant and positive correlations were found between PST and rate of plant production traits. In contrast, the correlations with PGT were negative. SY presented a significant and low positive correlation value with FT in the MA and AM RILs ($r = 0.16$, $r = 0.15$, $P \leq 0.01$). The correlations between SY and PGT, and PST were not significant except for PST in the MA RIL ($r = 0.37$, $P \leq 0.001$).

Marker Segregation Analysis and Consensus Genetic Map Construction

Six hundred and thirty-four markers were screened for DNA polymorphisms in the parents of MA RIL, which rendered a 36% polymorphism rate. Sixty-two (27%) out of the 228 polymorphic markers evaluated in the MA RIL population exhibited segregation distortion and thus, they could not be mapped. Finally, the genetic map of the MA RIL population (Figure not shown) consisted of 166 loci (158 SSRs, 1 SCAR, 6 SNPs and the *FIN* locus) distributed in 11 LGs. Out of 166 markers, 56 and 110 were dominant and codominant, respectively. LGs were named as reported in Pedrosa-Harand et al. (2008) using for the assignment of LG number and orientation 55 common SSR markers previously mapped (Freyre et al., 1998; Blair et al., 2008, 2010, 2011; Cichy et al., 2009; García et al., 2011). The map covered a genetic distance of 1188.9 cM, with an average of 7.2 cM among markers, which ranged from 5.3 cM (Pv02) to 16.0 cM (Pv05). The average genetic distance per LG was 108.1 cM, ranging from 70.6 cM (Pv10) to 134.1 cM (Pv07). A detailed description of this map is provided in Supplementary Table 5.

Likewise, 364 markers were screened for DNA polymorphisms in the parents of AM RIL, which rendered a 39% polymorphism rate. A total of 245 polymorphic markers were evaluated in the AM RIL population. Sixty-five (26%) out of these 245 markers, could not be mapped as they were not linked to other markers. Thus, the AM genetic map was constructed with a total of 180 loci (179 SSRs and *FIN* locus), from which 73 were dominant and 107 codominant. These loci were distributed among 11 LGs (Figure not shown) that covered a genetic distance of 1175.5 cM. The density of markers ranged from 3.4 cM (Pv01) to 12.6 cM (Pv04), with an average of 6.9 cM per LG. The longest LG was Pv08 (144.4 cM), whereas Pv11 was the shortest LG (48.2 cM),

with an average genetic distance of 106.9 cM per LG. A complete description of the AM map is shown in Supplementary Table 6.

The construction of a consensus map (Figure 2) was performed by connecting the MA and AM mapping data. To integrate both MA and AM maps in a single consensus map, 103 common markers were used as anchor points. As a result, marker segregation data were assembled for a total of 202 marker loci (196 SSRs, 5 SNPs, and the *FIN* locus) into 11 LGs. The total length of the consensus genetic map was 1156.2 cM and had a marker average density of one marker per 6.1 cM (Table 1). The marker order of the integrated map was largely collinear with the two individual maps, although a few local inversions and marker rearrangements over short intervals were observed. Most of the markers from both RIL populations showed a good linear relationship between their position on the genetic map and on the physical map of the common bean genome (Supplementary Table 7).

Multiple Environment Single-Locus QTL Analysis

QTL analysis based on MCIM mapping using QTLNetwork 2.0 was undertaken to identify single-locus QTLs across all environments. The positions of QTLs and their confidence intervals for the traits on the consensus map are shown in Figure 2 and Tables 2, 3. Thus, multi-environment QTL analyses allowed for the detection of 43 and 40 single-locus QTLs with significant additive main effects and/or QE effects for the evaluated quantitative traits in MA and AM RILs, respectively.

The distribution of the flowering and maturity time QTLs in the MA RIL population (Table 2) varied from 3 on Pv01 (*FT-1^{MA}*, *PGT-1^{MA}*, *PST-1^{MA}*), to 2 on each Pv02 (*PGT-2^{MA}*, *PST-2^{MA}*), and Pv09 (*PGT-9^{MA}*, *PST-9^{MA}*). All these QTLs had a positive additive value, indicating that alleles from the PHA0419 parent increase flowering and maturity times, whose main additive effects accounting for 3.84 (*PGT-9^{MA}*) to 18.96% (*FT-1^{MA}*) of the phenotypic variance for these traits. Five out of the seven QTLs identified showed QE interactions effects, ranged from 0.64 (for *PGT-2^{MA}* in G108) to 2.79% (for *PGT-1^{MA}* in F108). In addition, for the AM RIL population (Table 3), six single-locus QTLs were detected, three on Pv01 (*FT-1^{AM}*, *PGT-1^{AM}*, *PST-1^{AM}*), two on Pv05 (*FT-5^{AM}*, *PST-5^{AM}*), and one on Pv03 (*PGT-3^{AM}*). The additive effects of these QTLs explained a phenotypic variance up to 30.74% (*FT-1^{AM}*). Three of them exhibited significant QE interaction effects, which varied from 1.43 (for *PST-1^{AM}* in F109) to 18.89% (for *PST-1^{AM}* in G108). Most of the QTLs had a negative additive value, except for the QTL *PGT-3^{AM}*, which indicated that alleles from the PHA0399 parent mainly enhance flowering and maturity times. Combining both QTL mapping results, two genomic regions stood out, BMD045-*FIN* and *FIN*-BMC224 in MA and AM RIL populations, respectively, which contained QTLs controlling FT, PGT, and PST traits that were correlated in both mapping populations (Supplementary Table 4).

For vegetative growth habit traits, six single-locus QTLs were detected in the MA RIL population (Table 2), three located on Pv01 (*LMS-1^{MA}*, *NPB-1^{MA}*, *LI-1^{MA}*), two on Pv06 (*LMS-6^{MA}*,

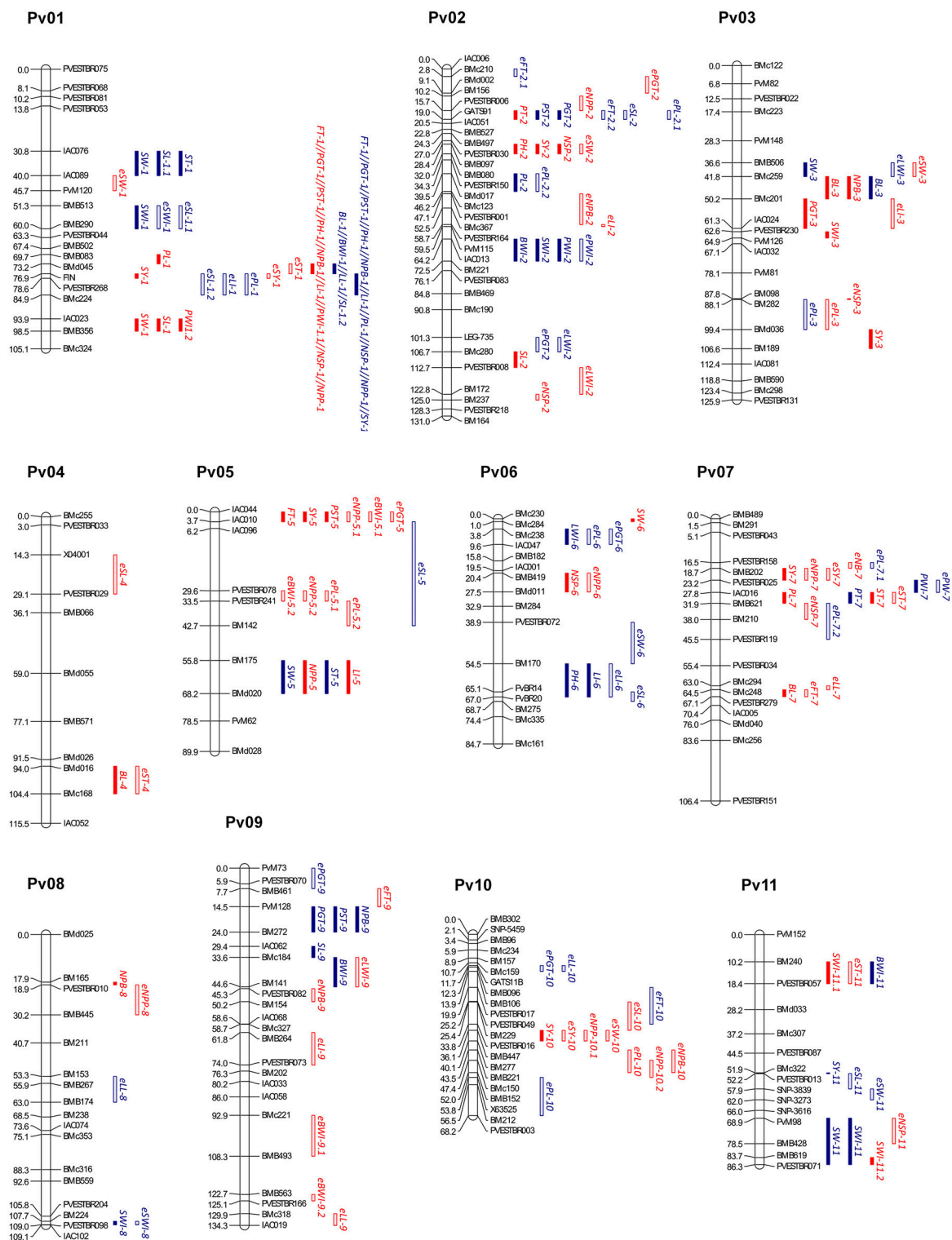


FIGURE 2 | Common bean consensus map showing the location of main-effect QTLs and E-QTLs for quantitative traits from the MA (blue color) and AM (red color) RIL populations. Distances among markers are indicated in cM to the left of the LGs. Names of markers are shown on the right. QTLs are depicted as vertical bars to the right of the LG. Names of QTLs are listed in **Tables 2–5**. Main effect QTLs are indicated with solid bars and E-QTLs are indicated with hatched bars.

TABLE 1 | Consensus map constructed from MA and AM RIL maps.

Linkage groups	Map length (cM)	No. of markers	Marker density (cM/marker)	Marker types		
				SSR	SNP	FIN
1	105.1	19	5.5	18	–	1
2	131	31	4.2	30	1	–
3	125.8	21	6	21	–	–
4	115.4	11	10.5	11	–	–
5	89.9	10	9	10	–	–
6	84.7	16	5.3	16	–	–
7	106.4	18	5.9	18	–	–
8	109.1	17	6.4	17	–	–
9	134.3	23	5.8	23	–	–
10	68.2	21	3.2	20	1	–
11	86.3	15	5.7	12	3	–
Total	1156.2	202	6.1	196	5	1

LI-6^{MA}), and one on Pv09 (*NPB-9^{MA}*), whose additive effects explained a total phenotypic variance that ranged from 3.29 (*LI-6^{MA}*) to 66.08% (*LMS-1^{MA}*). Two out of the six QTLs showed QE interaction effects, which ranged from 1.09 to 6.37% (for *LI-6^{MA}* in F108 and for *LI-1^{MA}* in G108, respectively). Regarding the AM RIL population, seven single-locus QTLs were identified (Table 3), three on Pv01 (*LMS-1^{AM}*, *NPB-1^{AM}*, *LI-1^{AM}*), and one on each Pv02 (*LMS-2^{AM}*), Pv03 (*NPB-3^{AM}*), Pv05 (*LI-5^{AM}*), and Pv08 (*NPB-8^{AM}*). The additive effects of these QTLs accounting for 1.55 (*LI-5^{AM}*) to 50.27% (*LMS-1^{AM}*); and four of them showed QE interaction effects that ranged from 2.57 to 9.84% (for *LI-1^{AM}* in F109 and G109, respectively). In both RIL populations, the same genomic region on Pv01 (*FIN-BMC224* and *BMD045-FIN*) contained QTLs for LMS, NPB, and LI traits. On this region, Mesoamerican alleles contributed to enhance the LMS and LI traits, whereas Andean alleles increased the NPB trait (Tables 2, 3).

Regarding plant production traits, 30 single-locus QTLs were detected in the MA RIL population (Table 2), 12 located on Pv01 (*BL-1^{MA}*, *BWI-1^{MA}*, *LL-1.1^{MA}*, *PL-1^{MA}*, *SL-1.1^{MA}*, *SL-1.2^{MA}*, *SWI-1^{MA}*, *ST-1^{MA}*, *SW-1^{MA}*, *NSP-1^{MA}*, *NPP-1^{MA}*, *SY-1^{MA}*), four on each Pv02 (*BWI-2^{MA}*, *PL-2^{MA}*, *PWI-2^{MA}*, *SWI-2^{MA}*) and Pv11 (*BWI-11^{MA}*, *SWI-11^{MA}*, *SW11^{MA}*, *SY-11^{MA}*), two on each Pv03 (*BL-3^{MA}*, *SW-3^{MA}*), Pv05 (*ST-5^{MA}*, *SW-5^{MA}*), Pv07 (*PWI-7^{MA}*, *PT-7^{MA}*), and Pv09 (*BWI-9^{MA}*, *SL-9^{MA}*), as well as one on each Pv06 (*LWI-6^{MA}*) and Pv08 (*SWI-8^{MA}*). Their additive effects accounting for a phenotypic variance that ranged from 1.77 (*BWI-9^{MA}*) to 18.90% (*SY-1^{MA}*). Ten out of the 30 single-locus QTLs displayed environment interactions effects that ranged from 0.51 to 2.71% (for *SWI-2^{MA}* and *LL-1.1^{MA}* in F108 and G108, respectively). Additionally, for the AM RIL population (Table 3), 27 single-locus QTLs were identified, eight located on Pv01 (*PL-1^{AM}*, *PWI-1.1^{AM}*, *PWI-1.2^{AM}*, *SL-1^{AM}*, *SW-1^{AM}*, *NSP-1^{AM}*, *NPP-1^{AM}*, *SY-1^{AM}*), four on each Pv02 (*PT-2^{AM}*, *SL-2^{AM}*, *NSP-2^{AM}*, *SY-2^{AM}*) and Pv07 (*BL-7^{AM}*, *PL-7^{AM}*, *ST-7^{AM}*, *SY-7^{AM}*), three on Pv03 (*BL-3^{AM}*, *SWI-3^{AM}*, *SY-3^{AM}*), two on each

Pv05 (*NPP-5^{AM}*, *SY-5^{AM}*), Pv06 (*SW-6^{AM}*, *NSP-6^{AM}*), and Pv11 (*SWI-11.1^{AM}*, *SWI-11.2^{AM}*), and one on each Pv04 (*BL-4^{AM}*) and Pv10 (*SY-10^{AM}*). The additive effects of these QTLs accounting for 1.32 (*SY-2^{AM}*) to 24.75% (*SL-1^{AM}*). Fourteen out of the 27 single-locus QTLs displayed environment interactions, whose effects explaining a phenotypic variance up to 14.71% (*NPP-1^{AM}* in F109). Furthermore, in MA and AM RIL populations, positive, and negative additive values were identified, indicating that alleles from both Mesoamerican and Andean parents have a positive agronomical effect on the crop growth and productivity traits.

Epistatic QTL Interactions

Epistatic and environment interactions among QTLs were detected by means of a two-dimensional genome scan using QTLNetwork 2.0. Thus, a total of 33 and 49 significant E-QTLs involved in 17 and 25 epistatic interactions were identified for the MA and AM RIL populations, respectively (Tables 4, 5). Most of these interactions were due to loci without detectable QTL additive main-effects, and only five and three E-QTLs had both individual additive and epistatic effects in the MA and AM RIL populations, respectively. No significant epistatic interactions were detected for PST, LMS, BL, and PT traits in both RIL populations; whereas E-QTLs were not identified for NPB, ST, NSP, NPP, and SY, and for PWI and SWI traits in the MA and AM RIL population, respectively (Tables 4, 5).

For flowering and maturity time traits, four epistatic interactions were identified in the MA RIL population (Table 4) that explained a phenotypic variance that ranged from 1.09 (*ePGT-2^{MA}* × *ePGT-10^{MA}*) to 3.82% (*eFT-2.1^{MA}* × *eFT-10^{MA}*). For the AM RIL population (Table 5), two epistatic interactions were detected whose effects accounting for up to 5.73% (*eFT-7^{AM}* × *eFT-9^{AM}*) of the phenotypic variance. Among the six epistatic interactions detected for both RIL populations, three showed significant environmental interactions (E-QE) effects that ranged from 0.72 (*ePGT-2^{MA}* × *ePGT-10^{MA}* in G109) to 6.09% (*eFT-7^{AM}* × *eFT-9^{AM}* in F109).

Regarding vegetative growth traits, only one epistatic interaction was identified in the MA RIL population (Table 4), which did not show E-QE effects and explained the 2.34% of the phenotypic variance for the LI trait; whereas four epistatic interactions were detected in the AM RIL population (Table 5) whose effect ranged from 1.16 (*eLI-2^{AM}* × *eLI-3^{AM}*) to 4.73% (*eNPB-2^{AM}* × *eNPB-10^{AM}*). These four epistatic interactions displayed E-QE effects that explained up to 5.32% (*eNPB-7^{AM}* × *eNPB-9^{AM}* in F108) of the phenotypic variance.

For plant production traits, 12 epistatic interactions were identified in the MA RIL population (Table 4) and their additive by additive epistatic effects ranged from 1.11 (*ePWI-2^{MA}* × *ePWI-7^{MA}*) to 8.30% (*eSL-1.1^{MA}* × *eSL-2^{MA}*). Four out of the 12 epistatic interactions showed significant environment interaction effects that explained up to 1.29% (*eSW-6^{MA}* × *eSW-11^{MA}* in F109) of the phenotypic variance. For the AM RIL population (Table 5), 19 epistatic interactions were detected whose effects ranged from 1.51 (*eST-7^{AM}* × *eST-11^{AM}*) to 12.20% (*eNPP-2^{AM}* × *eNPP-5.2^{AM}*). Ten out of the 19 epistatic interactions displayed significant environment interactions and their effects

TABLE 2 | Single-locus QTLs and QTLs × Environment (QE) interaction effects identified in the MA RIL population grown in four different environments.

QTL	Marker interval	LG (pos.) ^a	F-value ^b	A ^c	R ² (a) ^d	QE AE ^e			R ² (ae) ^f
FLOWERING AND MATURITY TRAITS									
FT-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	17.33	1.43**	18.96		ns		
PGT-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	21.26	1.53**	10.26	–1.34	**	F108	2.79
						1.20	**	G108	1.90
						0.86	**	G109	0.95
PGT-2 ^{MA}	PVESTBR006-GATs91	Pv02 (15.7–19.0)	8.28	0.91**	4.24	–0.72	*	F108	0.78
						0.81	*	G108	0.64
						0.94	**	G109	0.87
						–1.01	*	F109	1.04
PGT-9 ^{MA}	PVM128-BM272	Pv09 (14.5–24.0)	7.42	1.16**	3.84	1.1	**	G109	1.14
PST-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	12.73	1.66**	7.62	0.91	**	G108	1.39
PST-2 ^{MA}	PVESTBR006-GATs91	Pv02 (15.7–19.0)	8.45	2.08**	6.69		ns		
PST-9 ^{MA}	PVM128-BM272	Pv09 (14.5–24.0)	6.75	1.48**	4.29	1.04	*	G109	1.33
VEGETATIVE GROWTH TRAITS									
LMS-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	181.35	0.56**	66.08		ns		
LMS-6 ^{MA}	BM170-PVBR20	Pv06 (54.5–66.9)	9.23	0.12**	5.56		ns		
NPB-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	12.73	–0.13**	7.79		ns		
NPB-9 ^{MA}	PVM128-BM272	Pv09 (14.5–24.0)	9.17	0.11**	6.27		ns		
LJ-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	12.05	0.04**	16.41	0.13	**	G108	6.37
						–0.06	*	F109	1.42
						–0.05	*	F108	1.09
LJ-6 ^{MA}	BM170-PVBR20	Pv06 (54.5–66.9)	7.79	0.05**	3.29	0.08	**	G108	3.12
PLANT PRODUCTION TRAITS									
BL-1 ^{MA}	BMD045-FIN	Pv01 (73.2–76.9)	7.22	0.03**	3.09	0.05	**	G108	2.12
						–0.04	**	F109	1.86
BL-3 ^{MA}	BMC259-BMC201	Pv03 (41.8–50.2)	7.8	–0.02**	13.17	–0.03	**	G108	1.33
						0.03	*	F109	1.64
BWI-1 ^{MA}	BMD045-FIN	Pv01 (73.2–76.9)	7.05	0.03**	3.06		ns		
BWI-2 ^{MA}	IAC013-BM221	Pv02 (64.2–72.5)	6.93	–0.04**	4.02		ns		
BWI-9 ^{MA}	BMC184-BM141	Pv09 (33.6–44.6)	7.38	–0.02**	1.77		ns		
BWI-11 ^{MA}	BM240-PVESTBR057	Pv11(10.2–18.4)	7.17	–0.03**	3.31		ns		
LL-1.1 ^{MA}	BMD045-FIN	Pv01 (73.2–76.9)	7.05	0.01**	2.85	0.02	**	G108	2.71
						–0.01	**	G109	1.51
						–0.01	**	F109	1.54
LWI-6 ^{MA}	BMc238-IAC047	Pv06 (3.8–9.6)	6.73	–0.16**	3.70		ns		
PL-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	7.07	0.39**	7.59		ns		
PL-2 ^{MA}	BMD017-BMC123	Pv02 (39.5–46.2)	10.81	–0.40**	4.75		ns		
PWI-2 ^{MA}	IAC013-BM221	Pv02 (64.2–72.5)	7.1	–0.24**	2.56	0.21	*	F108	1.86
PWI-7 ^{MA}	PVESTBR025-IAC016	Pv07 (23.2–27.8)	10.85	–0.34**	3.68	0.26	*	F108	1.81
PT-7 ^{MA}	IAC016- BMB621	Pv07 (27.8–31.9)	7.22	–0.20**	2.88	0.25	*	F108	2.07
SL-1.1 ^{MA}	IAC076-IAC089	Pv01 (30.8–40.0)	7.86	0.34**	4.96		ns		
SL-1.2 ^{MA}	BMD045-FIN	Pv01 (73.2–76.9)	10.09	0.40**	4.82		ns		
SL-9 ^{MA}	IAC062-BMC184	Pv09 (29.4–33.6)	7.91	–0.44*	9.65		ns		
SWI-1 ^{MA}	BMB513-BMB290	Pv01 (51.3–60.0)	7.85	0.26**	5.23		ns		
SWI-2 ^{MA}	IAC013-BM221	Pv02 (64.2–72.5)	7.22	–0.16**	3.47	0.11	*	F108	0.51
SWI-8 ^{MA}	BM224-PVESTBR098	Pv08 (107.6–109.0)	8.11	0.24**	3.03		ns		
SWI-11 ^{MA}	PVM98-PVESTBR071	Pv11 (68.9–86.3)	7.93	–0.20**	3.34		ns		
ST-1 ^{MA}	IAC076-IAC089	Pv01 (30.8–40.0)	6.96	0.09**	1.92		ns		
ST-5 ^{MA}	BM175-BMD020	Pv05 (55.8–68.2)	6.7	–0.13**	4.40		ns		
SW-1 ^{MA}	IAC076-IAC089	Pv01 (30.8–40.0)	6	0.06**	4.87		ns		
SW-3 ^{MA}	BMB506-BMC259	Pv03 (36.6–41.8)	9.21	0.07**	4.38		ns		

(Continued)

TABLE 2 | Continued

QTL	Marker interval	LG (pos.) ^a	F-value ^b	A ^c	R ² (a) ^d	QE AE ^e	R ² (ae) ^f
SW-5 ^{MA}	BM175-BMD020	Pv05 (55.8–68.2)	6.61	−0.05**	5.12	ns	
SW11 ^{MA}	PVM98-PVESTBR071	Pv11 (68.9–86.3)	6.29	−0.05**	4.81	ns	
NSP-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	13.21	0.24**	7.90	−0.15 **	F108 1.65
						0.18 **	G108 2.19
NPP-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	17.76	0.48**	9.51	−0.29 **	F108 2.17
						0.36 **	G108 1.92
						0.2 *	G109 0.76
SY-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	32.15	0.37**	18.90	−0.17 **	F108 1.91
						0.12 *	G108 0.84
SY-11 ^{MA}	BMC322-PVESTBR013	Pv11 (51.9–52.2)	8.37	0.23**	6.12	ns	

^aLinkage group and the estimated confidence interval of QTL position in the consensus map (in Kosambi cM).

^bF-values of significance of each QTL.

^cEstimated additive effect. Positive values indicate that alleles from PHA0419 have a positive effect on the traits, and negative values indicate that positive effect on the traits is due to the presence of the alleles from BELUGA.

^dPercentage of the phenotypic variation explained by additive effects.

^ePredicted additive by environment interaction effect. The meaning of sign values is described in the second footnote (^c).

^fPercentage of the phenotypic variation explained by additive by environment interaction effect.

*P ≤ 0.05, **P ≤ 0.01. Experiment wide P-value. Only significant effects are listed. ns, not significant effects on the four environments evaluated.

FT, flowering time; PGT, pod green time; PST, physiological maturity time; LMS, length of main stem; NPB, the number of primary stem branches; LI, internode length; BL, bracteole length; BWI, bracteole width; LL, leaflet length; LWI, leaflet width; PL, pod length; PWI, pod width; PT, pod thickness; SL, seed length; SWI, seed width; ST, seed thickness; SW, 100 seed weight; NSP, number of seeds per pod; NPP, number of pods per plant; SY, seed yield.

ranged from 1.18 ($eSY-7^{AM} \times eSY-10^{AM}$ in G108) to 12.45% ($eNPP-5.1^{AM} \times eNPP-8^{AM}$ in F109).

DISCUSSION

Many traits of agronomic or biological importance undergo dynamic phenotypic changes during vegetative growth. In fact, the temporal control of flowering initiation determines the time invested in vegetative growth and, consequently, the vegetative resources available during reproduction. In addition, the rate of leaf, flower, and fruit production is a major component of vegetative growth. There is a close association between flowering initiation and vegetative growth; however, how these processes are coordinated during plant development remains unexplored. In this study, the variation for vegetative growth, the rate of leaf, flower, and fruit production, alongside their relationships with flowering and fruit time were investigated in two RIL populations from inter-gene pool crosses of an indeterminate Mesoamerican race Durango and a determinate Andean cultivar race Nueva Granada with the aim to unravel the genetic dynamics underlying vegetative growth and time to flowering.

The Genetic Architecture of Vegetative Growth and Flowering Time in Common Bean

A prerequisite for QTL mapping is the assessment of the quantitative trait in multiple environments. In this study, agronomic evaluation was assessed in four different environments, under open field and greenhouse conditions. The analysis of the two RIL populations under these environments

indicates predominantly quantitative inheritance rather than qualitative genes controlling vegetative growth and time to flowering. However, not only additive main effects are responsible for the phenotypic variation observed in our RIL mapping populations, but also epistatic interactions play an important role on the genetic control of flowering time and rate of plant production traits.

In both MA and AM RIL populations, normal distributions were found for most traits, although it was detected that the length of main stem could be controlled by a single gene in the AM RIL population since a bimodal distribution was found in the determinate type I Andean × indeterminate type IV Mesoamerican cross. This bimodal trend is adjusted with the discrete classes and the proportion expected for an autosomal major gene. Simple and complex genetic model has previously been proposed for plant height in common bean. Thus, Kornegay et al. (1992) observed that plant height is a trait of simple inheritance and high heritability. However, Frazier et al. (1958) stated that to reach again the erect trait in a plant of typically determinate growth habit, in addition to the *FIN* locus, it is needed the action of at least three recessive genes, or a set of minor action genes. Likewise, Davis and Frazier (1966) predicted several genes for internode length, as well as Checa et al. (2006) in indeterminate/indeterminate crosses of Andean and Mesoamerican beans; these authors found that the inheritance of plant height and internode length was mostly additive with only a few genes involved in the expression, and that these genes were most likely modified by interaction with minor genes and with the environment. Checa and Blair (2008) observed a quantitative inheritance rather than qualitative genes controlling plant height in an indeterminate type IV Mesoamerican × indeterminate type II Andean cross, although they also detected a relatively major

TABLE 3 | Single-locus QTLs and QTLs × Environment (QE) interaction effects identified in the AM RIL population grown in four different environments.

QTL	Marker interval	LG (pos.) ^a	F-value ^b	A ^c	R ² (a) ^d	QE AE ^e		R ² (ae) ^f
FLOWERING AND MATURITY TRAITS								
FT-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	21.37	−6.82**	30.74		ns	
FT-5 ^{AM}	IAC044-IAC010	Pv05 (0.0–3.7)	5.91	−1.99**	10.70		ns	
PGT-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	16.49	−5.54**	21.29	−4.59	**	G108 4.96
PGT-3 ^{AM}	BMC201-IAC024	Pv03 (50.2–61.3)	5.4	1.61**	3.82	2.13	**	F109 1.96
PST-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	25.07	−8.94**	31.91	6.06	*	F109 1.43
						−15.2	**	G108 18.89
PST-5 ^{AM}	IAC044-IAC010	Pv05 (0.0–3.7)	6.27	−2.29**	9.13		ns	
VEGETATIVE GROWTH TRAITS								
LMS-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	33.98	−170.17**	50.27	97.4	**	F109 9.51
LMS-2 ^{AM}	BMB097-BMB080	Pv02 (28.4–32.0)	8.8	55.33**	5.97	76.8	**	F109 4.25
NPB-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	14.85	0.15**	9.23		ns	
NPB-3 ^{AM}	BMC259-BMC201	Pv03 (41.8–50.2)	8.97	−0.02*	2.95	0.08	*	F109 3.11
NPB-8 ^{AM}	BM165-PVESTBR010	Pv08 (17.8–18.8)	10.75	−0.05*	5.29		ns	
LJ-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	107.31	−4.46**	34.28	4.4	**	F108 9.24
						2.43	**	F109 2.57
						−4.07	**	G108 6.6
						−6.78	**	G109 9.84
LJ-5 ^{AM}	BM175-BMD020	Pv05 (55.8–68.2)	5.4	0.99**	1.55		ns	
PLANT PRODUCTION TRAITS								
BL-3 ^{AM}	BMC259-BMC201	Pv03 (41.8–50.2)	9.28	0.10**	17.32		ns	
BL-4 ^{AM}	BMD016-BMC168	Pv04 (94.0–104.4)	5.77	0.09**	2.45		ns	
BL-7 ^{AM}	BMC248-PVESTBR279	Pv07 (64.5–67.1)	8.17	−0.04**	7.15		ns	
PL-1 ^{AM}	BMB083-BMD045	Pv01 (69.7–73.2)	5.8	0.24**	3.98	−0.18	**	F108 3.85
PL-7 ^{AM}	IAC016-BMB621	Pv07 (27.8–31.9)	8.73	0.56**	10.37	−0.15	**	F108 3.66
PWI-1.1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	7.11	−0.65**	8.74	0.53	*	F108 2.91
PWI-1.2 ^{AM}	IAC023-BMB356	Pv01 (93.9–98.5)	6.62	0.83**	6.55		ns	
PT-2 ^{AM}	PVESTBR006-GATS91	Pv02 (15.7–19.0)	6.36	0.25**	5.60		ns	
SL-1 ^{AM}	IAC023-BMB356	Pv01 (93.9–98.5)	9.05	1.88**	24.75		ns	
SL-2 ^{AM}	BMC280-PVESTBR008	Pv02 (106.7–112.7)	5.1	−0.23**	5.65		ns	
SWI-3 ^{AM}	PVESTBR230-BMB339	Pv03 (62.5–64.9)	10.36	−0.27**	10.95		ns	
SWI-11.1 ^{AM}	BM240-PVESTBR057	Pv11 (10.2–18.4)	9.1	−0.22**	13.29		ns	
SWI-11.2 ^{AM}	BMB619-PVESTBR071	Pv11 (83.7–86.3)	8.32	−0.49**	17.64		ns	
ST-7 ^{AM}	IAC016-BMB621	Pv07 (27.8–31.9)	5.89	0.18**	11.08		ns	
SW-1 ^{AM}	IAC023-BMB356	Pv01 (93.9–98.5)	5.98	0.13**	10.91		ns	
SW-6 ^{AM}	BMC230-BMC284	Pv06 (0.0–1.0)	5.59	0.04**	3.80		ns	
NSP-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	14.76	−0.46**	14.01	0.31	**	F108 2.09
						−0.62	**	F109 9.65
NSP-2 ^{AM}	BMB097-BMB080	Pv02 (28.4–32.0)	6.06	0.32**	6.39	0.3	**	F109 3.24
NSP-6 ^{AM}	BMB419-BMD011	Pv06 (20.4–27.5)	6.96	−0.37**	5.91	0.24	*	F108 1.15
						−0.54	**	F109 6.79
NPP-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	18.04	−0.82**	10.14	0.69	**	F108 2.73
						−1.64	**	F109 14.71
NPP-5 ^{AM}	BM175-BMD020	Pv05 (55.8–68.2)	6.20	0.42**	4.10	−0.64	**	F109 4.37
SY-1 ^{AM}	FIN-PVESTBR268	Pv01 (76.9–78.6)	9.17	−8.44**	14.75	−7.72	**	F109 6.55
SY-2 ^{AM}	BMB097-BMB080	Pv02 (28.4–32.0)	9.26	6.03**	1.32	6.06	**	F108 0.10
						9.17	**	F109 2.42
						5.97	**	G108 0.14
SY-3 ^{AM}	BMD036-BM189	Pv03 (99.4–106.6)	5.65	−2.09**	2.66	2.09	*	G108 0.68

(Continued)

TABLE 3 | Continued

QTL	Marker interval	LG (pos.) ^a	F-value ^b	A ^c	R ² (a) ^d	QE AE ^e		R ² (ae) ^f	
SY-5 ^{AM}	IAC044-IAC010	Pv05 (0.0–3.7)	5.71	0.86*	4.48	−5.73	**	F109	3.57
SY-7 ^{AM}	BMB202-PVESTBR025	Pv07 (18.7–23.2)	6.12	−3.59**	7.10	3.34	**	F108	0.39
						3.44	**	G108	0.85
SY-10 ^{AM}	BMB447-BM277	Pv10 (36.1–40.1)	10.31	0.20*	1.80	4.05	**	F109	1.30

^aLinkage group and the estimated confidence interval of QTL position in the consensus map (in Kosambi cM).

^bF-values of significance of each QTL.

^cEstimated additive effect. Positive values indicate that alleles from BELUGA have a positive effect on the traits, and negative values indicate that positive effect on the traits is due to the presence of the alleles from PHA0399.

^dPercentage of the phenotypic variation explained by additive effects.

^ePredicted additive by environment interaction effect. The meaning of sign values is described in the second footnote (^c).

^fPercentage of the phenotypic variation explained by additive by environment interaction effect.

* $P \leq 0.05$, ** $P \leq 0.01$. Experiment wide P-value. Only significant effects are listed. ns, not significant effects on the four environments evaluated.

FT, flowering time; PGT, pod green time; PST, physiological maturity time; LMS, length of main stem; NPB, the number of primary stem branches; LI, internode length; BL, bracteole length; BWI, bracteole width; LL, leaflet length; LWI, leaflet width; PL, pod length; PWI, pod width; PT, pod thickness; SL, seed length; SWI, seed width; ST, seed thickness; SW, 100 seed weight; NSP, number of seeds per pod; NPP, number of pods per plant; SY, seed yield.

gene or single locus with pleiotropic effects on plant height and internode length.

Transgressive segregation in both directions was observed for most of the evaluated traits in both RIL populations, except for the tendency of most lines to present averages for LMS and LI traits in the MA RIL population, and for SY and NPP traits in both RIL populations closer to the determinate Andean genotype, which might be related to segregation distortion. Thus, positive and negative transgressive segregations observed suggest that parental lines bear alleles that contribute to vegetative growth and time to flowering variation at several loci. High and significant heritability estimates, as well as positive correlations between LMS with FT and rate of plant production were found. FT was associated with initiation to immature or green pod and physiological pod or dry pod. In spite of the fact that FT and PST correlated with rate of plant production traits, it was not found correlation with PGT. Tar'an et al. (2002) also revealed positive correlations between components of plant height with FT and rate of seed production. Likewise, Koinange et al. (1996) reported that the gene *FIN* controlling determinacy has pleiotropic effects on FT, PST, and rate of plant production. In this work, it is shown that there is a cause-effect relation among LMS, FT, and rate of plant production traits, suggesting that physically linked or pleiotropic genes might be involved in the regulation of these traits (Aastveit and Aastveit, 1993).

Most of the single-locus QTLs detected in this work overlap with QTLs identified in some quantitative analyses of flowering time and vegetative growth carried out in common bean (Blair et al., 2006, 2010; Checa and Blair, 2008; Wright and Kelly, 2011). In general, these single-locus QTLs were responsible for the majority of the genetic variation for rate of plant production and time to flowering traits within common bean populations. However, it should be noted the presence of a considerable number of epistatic interactions for vegetative growth, plant production and flowering time traits that have not been reported so far. Taken together, these epistatic interactions reinforce the hypothesis that epistasis is involved in the genetic control of agronomical traits as well as epistatic interactions are more

frequent in inter-gene pool crosses than in intra-gene pool crosses of common bean (Borel et al., 2016). Thus, for example, significant genetic interactions were found for the genomic region where the *FIN* locus is located (Pv01) with other genomic regions on Pv02, Pv06, and Pv11 in the MA RIL population, as well as on Pv04 in the AM RIL population. Epistatic interactions for flowering time traits were also detected in other genomic regions, whose effects explained up to 5.73% (*eFT-7^{AM}* and *eFT-9^{AM}* on Pv07 and Pv09, respectively) of the phenotypic variance.

Seed yield is mainly determined by factors such as number of seeds per pod and seed weight. In this study, the skewness to the lower values shown by the two related-productivity traits, SY and NPP, is in agreement with previous studies (Singh and Urrea, 1995; Johnson and Gepts, 1999; Bruzi et al., 2007), where the biological constraints of the inter-gene pool crosses, Andean and Mesoamerican, hamper reaching the maximum possible yields. Said limitation might result from the loss of favorable epistatic combinations or low probability of recovering superior genotype combination, among other reasons (Johnson and Gepts, 2002; Moreto et al., 2012). According to this hypothesis, the high number of epistatic interactions detected for all yield components is remarkable. Thus, epistatic interaction effects accounted for more than 10% of the phenotypic variance for PL (12.54%) and SL (13.7%) traits in the MA RIL population, as well as for ST (17.29%), SW (15.4%), and NSP (14.63%) traits in the AM RIL population. Hence, results of this research showed the importance of the epistatic effects in the genetic regulation of yield component traits. Thereby, both main and epistatic interaction effects should be considered for a successful application of marker assisted selection (MAS) programs in order to increase yield in common bean.

Rate of Plant Production and Time to Flowering Genetic Links

In order to determine the genetic basis of the rate of plant production during the vegetative growth and time to flowering, QTL mapping was carried out with the average traits estimated in different conditions from flower to seed components. In both

TABLE 4 | Epistatic QTLs (E-QTLs) and E-QTL x Environment (E-QE) interaction effects detected in the MA RIL populations grown in four different environments.

e-QTL ^a	Marker interval	LG (pos.) ^b	e-QTLj ^a	Marker interval	LG (pos.)	F-value ^c	AA ^d	R ² (aa) ^e	e-QE AAE ^f	R ² (aae) ^g
FLOWERING AND MATURITY TRAITS										
eFT-2.1 ^{MA}	IAC006-BMC210	Pv02 (0.00–2.8)	eFT-10 ^{MA}	PVESTBR017-PVESTBR016	Pv10 (19.9–33.8)	11.88	1.07**	3.82	1.27	**
eFT-2.2 ^{MA}	PVESTBR006-GATS91	Pv02 (15.7–19.0)	eFT-10 ^{MA}	PVESTBR017-PVESTBR016	Pv10 (19.9–33.8)	9.59	0.92**	2.47	ns	
ePGT-2 ^{MA}	LEG-735-Bmc280	Pv02 (101.3–106.4)	ePGT-10 ^{MA}	GATS11B-BMB106	Pv10 (11.7–13.9)	5.82	–1.56**	1.09	1.98	*
ePGT-6 ^{MA}	BMc238-IAC047	Pv06 (3.8–9.6)	ePGT-g ^{MA}	PVM73-BMB461	Pv09 (0.0–7.6)	5.61	–0.81**	1.61	–1.75	*
VEGETATIVE GROWTH TRAITS										
eLI-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	eLI-g ^{MA}	BM170-PVBR20	Pv06 (54.5–67.0)	5.89	–0.06**	2.34	ns	
PLANT PRODUCTION TRAITS										
eLL-8 ^{MA}	BM153-BMB174	Pv08 (63.3–63.0)	eLL-10 ^{MA}	GATS11B-BMB106	Pv10 (11.7–13.9)	7.89	–0.01**	3.69	ns	
eLWI-2 ^{MA}	LEG-735-BMC280	Pv02 (101.3–106.7)	eLWI-3 ^{MA}	BMB506-BMC259	Pv03 (36.6–41.8)	6.82	–0.27**	6.03	–0.13	*
ePL-1 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	ePL-2.2 ^{MA}	BMD017-BMC123	Pv02 (39.5–46.2)	5.75	–0.19*	1.14	0.26	*
ePL-2.1 ^{MA}	PVESTBR006-GATS91	Pv02 (15.7–18.9)	ePL-10 ^{MA}	XG3525-PVESTBR003	Pv10 (53.8–68.2)	9.61	–0.38**	3.70	ns	
ePL-3 ^{MA}	BMC282-BMD036	Pv03 (88.1–99.4)	ePL-7.1 ^{MA}	PVESTBR158-BMB202	Pv07 (16.5–18.7)	11.77	–0.58**	4.08	ns	
ePL-6 ^{MA}	BMC238-IAC047	Pv06 (3.8–9.6)	ePL-7.2 ^{MA}	BMB621-PVESTBR119	Pv07 (31.9–45.5)	9.65	0.41**	3.62	ns	
ePW1-2 ^{MA}	IAC013-BM221	Pv02 (64.2–72.5)	ePW-7 ^{MA}	PVESTBR025-IAC016	Pv07 (23.2–27.8)	8.14	0.21**	1.11	ns	
eSL-1.1 ^{MA}	BMB513-BMB290	Pv01 (51.3–59.9)	eSL-2 ^{MA}	PVESTBR006-GATS91	Pv02 (15.7–18.9)	7.34	–0.86**	8.30	ns	
eSL-1.2 ^{MA}	FIN-BMC224	Pv01 (76.9–84.9)	eSL-11 ^{MA}	PVESTBR013-SNP-3839	Pv11 (52.2–57.9)	7.3	–0.27**	1.51	ns	
eSL-5 ^{MA}	IAC010-BM142	Pv05 (3.7–42.7)	eSL-6 ^{MA}	PVBR14-BM275	Pv06 (65.1–68.7)	7.83	–0.57**	3.89	0.62	**
eSW1-1 ^{MA}	BMB513-BMB290	Pv01 (51.3–59.9)	eSW1-g ^{MA}	BM224-PVESTBR098	Pv08 (107.6–109.0)	6.64	0.23**	1.91	ns	
eSW-6 ^{MA}	PVESTBR072-BM170	Pv06 (88.9–54.5)	eSW-11 ^{MA}	SNP-3839-SNP-3273	Pv11 (57.9–62.0)	7.87	0.05**	2.83	–0.09	**

^aE-QTL and E-QTLj are the two QTLs involved in epistatic interaction.

^bLinkage group and the estimated confidence interval of QTL position in the consensus map (in Kosambi cM).

^cF-values of significance of each QTL.

^dEstimated additive by additive epistatic effect. Positive values indicate that alleles from PHA0419 have a positive effect on the traits, and negative values indicate that positive effect on the traits is due to the presence of the alleles from BELUGA.

^ePercentage of the phenotypic variation explained by additive by additive epistatic effects.

^fPredicted additive by additive epistatic effect by environment interaction effect. The meaning of sign values is described in the third footnote (^d).

^gPercentage of the phenotypic variation explained by additive by additive epistatic effect by environment interaction effect.

*P ≤ 0.05, **P ≤ 0.01. Experiment wide P-value. Only significant effects are listed, ns, not significant effects on the four environments evaluated.

FT: flowering time; PGT: pod green time; LI: internode length; LWI: leaflet width; PL: leaflet length; PVW: pod width; SL: seed length; SW: 100 seed weight.

TABLE 5 | Epistatic QTLs (E-QTLs) and E-QTL × Environment (E-QE) interaction effects detected in the AM RIL populations grown in four different environments.

e-QTL ^a	Marker interval	LG (pos.) ^b	e-QTL ^a	Marker interval	LG (pos.)	F-value ^c	AA ^d	R ² (aa) ^e	e-QE AAE ^f	R ² (aa) ^g
FLOWERING AND MATURITY TRAITS										
eFT-7 ^{AM}	BMC248-PVESTBR279	Pv07 (64.5–67.1)	eFT-9 ^{AM}	BMB461-PVM128	Pv09 (7.6–14.5)	8.08	–6.03**	5.73	4.16 *	F108 0.77
									–11.02 **	F109 6.09
									6.10 **	G103 1.61
ePGT-2 ^{AM}	BMC210-BMD002	Pv02 (2.8–9.1)	ePGT-5 ^{AM}	IAC044-IAC010	Pv05 (0.0–3.7)	5.83	1.63**	3.89	ns	
VEGETATIVE GROWTH TRAITS										
eNPB-2 ^{AM}	PVESTBR001-PVESTBR164	Pv02 (47.1–58.7)	eNPB-10 ^{AM}	BMB221-BMB152	Pv10 (43.5–52.0)	14.51	0.08**	4.73	0.07 *	F108 2.69
eNPB-7 ^{AM}	PVESTBR158-BMB202	Pv07 (16.4–18.6)	eNPB-9 ^{AM}	PVESTBR082-BM154	Pv09 (45.3–50.1)	11.07	–0.11**	3.87	–0.17 **	F108 5.32
eLL-2 ^{AM}	PVESTBR164-PVM115	Pv02 (58.7–59.5)	eLL-3 ^{AM}	BMC201-IAC024	Pv03 (50.2–61.3)	5.97	0.31*	1.16	–1.51 **	F109 1.51
eLL-3 ^{AM}	BMC201-IAC024	Pv03 (50.2–61.3)	eLL-9 ^{AM}	BMB264-PVESTBR073	Pv09 (61.8–73.9)	6.93	–1.03**	2.13	1.34 *	G108 0.76
									0.85 *	F108 0.60
									–0.96 *	G103 0.52
PLANT PRODUCTION TRAITS										
eBWI-5.1 ^{AM}	IAC044-IAC010	Pv05 (0.0–3.7)	eBWI-9.1 ^{AM}	BMC221-BMB493	Pv09 (92.9–108.3)	4.97	–0.09**	6.34	ns	
eBWI-5.2 ^{AM}	PVESTBR078-PVESTBR241	Pv05 (29.6–33.5)	eBWI-9.2 ^{AM}	BMB563-PVESTBR166	Pv09 (122.7–125.1)	8.52	–0.13**	10.82	ns	
eLL-7 ^{AM}	BMC294-BMC248	Pv07 (63.0–64.5)	eLL-9 ^{AM}	BMC318-IAC019	Pv09 (129.9–134.3)	5.04	–0.02**	4.64	–0.03 **	G109 4.94
eLWI-2 ^{AM}	PVESTBR008-BM172	Pv02 (112.7–122.8)	eLWI-9 ^{AM}	BMC184-BM141	Pv09 (33.6–44.6)	5.73	–0.02**	6.59	ns	
ePL-3 ^{AM}	BMD282-BMD036	Pv03 (88.1–99.4)	ePL-5.1 ^{AM}	PVESTBR078-PVESTBR241	Pv05 (29.6–33.5)	5.76	–0.77**	2.89	0.85 *	F108 2.90
ePL-5.2 ^{AM}	PVESTBR241-BM142	Pv05 (33.5–42.7)	ePL-10 ^{AM}	BMB221-BMB152	Pv10 (43.5–52.0)	5.46	0.40**	3.65	–0.37 *	F108 1.99
eSL-4 ^{AM}	X04001-PVESTBR029	Pv04 (14.3–29.1)	eSL-10 ^{AM}	BMB447-BM277	Pv10 (36.1–40.1)	7.32	0.66**	8.78	ns	
eST-1 ^{AM}	BMD045-FIN	Pv01 (73.2–76.9)	eST-4 ^{AM}	BMD016-BMC168	Pv04 (94.0–104.4)	7.83	0.13**	5.09	–0.13 *	G108 1.84
eST-7 ^{AM}	IAC016-BMB621	Pv07 (27.8–31.9)	eST-11 ^{AM}	BMD240-PVESTBR057	Pv11 (10.2–18.3)	6.97	0.19**	12.20	ns	
eSW-1 ^{AM}	IAC089-FM120	Pv01 (40.0–45.7)	eSW-10 ^{AM}	BMB447-BM277	Pv10 (36.1–40.1)	6.71	–0.23**	10.75	ns	
eSW-2 ^{AM}	BMB097-BMB080	Pv02 (28.4–32.0)	eSW-3 ^{AM}	BMB506-BMC259	Pv03 (36.6–41.8)	5.86	0.19**	4.65	ns	
eNSP-2 ^{AM}	BM172-BM237	Pv02 (122.8–125.0)	eNSP-3 ^{AM}	BMB098-BM282	Pv03 (87.7–88.1)	6.64	0.66**	10.10	–0.38 *	G108 1.88
eNSP-7 ^{AM}	BMB621-BM210	Pv07 (31.9–38.0)	eNSP-11 ^{AM}	PVM98-BMB428	Pv11 (68.9–78.5)	6.94	0.19**	4.53	ns	
eNPP-2 ^{AM}	BM156-PVESTBR006	Pv02 (10.2–15.7)	eNPP-5.2 ^{AM}	PVESTBR078-PVESTBR241	Pv05 (29.6–33.5)	9.01	–0.18*	1.51	–0.31 *	F109 1.18
eNPP-5.1 ^{AM}	IAC044-IAC010	Pv05 (0.0–3.7)	eNPP-8 ^{AM}	PVESTBR010-BMB445	Pv08 (18.8–30.2)	7.23	–0.39**	2.97	0.53 **	F108 1.82
									–1.29** *	F109 12.45
eNPP-6 ^{AM}	BMB419-BMD011	Pv06 (20.4–27.5)	eNPP-10.1 ^{AM}	BMB447-BM277	Pv10 (36.1–40.1)	11.94	–0.38**	2.27	–0.68 **	F109 3.41
eNPP-7 ^{AM}	BMB202-PVESTBR025	Pv07 (18.7–23.2)	eNPP-10.2 ^{AM}	BMC150-X63525	Pv10 (47.4–53.8)	10.09	0.53**	2.51	1.47 **	F109 8.28
									–0.59 *	G108 1.27
eSY-1 ^{AM}	FIN-PVESTBR268	Pv01 (76.9–78.6)	eSY-10 ^{AM}	BMB447-BM277	Pv10 (36.1–40.1)	5.7	2.50**	1.70	ns	
eSY-7 ^{AM}	BMB202-PVESTBR025	Pv07 (18.7–23.2)	eSY-10 ^{AM}	BMB447-BM277	Pv10 (36.1–40.1)	7.55	3.63**	4.69	–3.31 **	F108 1.10
									–3.55 **	G108 0.18

^aE-QTL and E-QTLj are the two QTLs involved in epistatic interaction.^bLinkage group and the estimated confidence interval of QTL position in the consensus map (in Kosambi cM).^cF-values of significance of each QTL.^dEstimated additive by additive epistatic effect. Positive values indicate that alleles from BELUGA have a positive effect on the traits, and negative values indicate that positive effect on the traits is due to the presence of the alleles from PHA0399.^ePercentage of the phenotypic variation explained by additive by additive epistatic effects.^fPredicted additive by additive epistatic effect by environment interaction effect. The meaning of sign values is described in the third footnote (^d).^gPercentage of the phenotypic variation explained by additive by additive epistatic effect by environment interaction effect.^h*P ≤ 0.05, **P ≤ 0.01. Experiment wide P-value. Only significant effects are listed; ns, not significant effects on the four environments evaluated.

FT, flowering time; PGT, pod green time; NPB, the number of primary stem branches; LI, internode length; BMB, bracteole width; LL, leaflet length; LWI, leaflet width; PL, pod length; SL, seed length; ST, seed thickness; SW, 100 seed weight; NSP, number of seeds per pod; NPP, number of pods per plant; SY, seed yield.

MA and AM RIL populations, it was detected a QTL located on Pv01 at the *FIN* locus, which showed large additive relative effects on time to flowering traits (up to 19% for FT and 32% for PST of the phenotypic variation in the MA and AM RIL populations, respectively). Comparative QTL analyses showed that this genomic region on Pv01 was also involved in the regulation of vegetative growth traits. Thus, QTLs for LMS, NPB, and LI traits were detected, which explained large additive effects (50–66% of the phenotypic variance for LMS). Within this genomic region, additional QTLs controlling pod size and productivity components (NSP and NPP) which explained more than 8% of the phenotypic variance. Furthermore, in both RIL populations, the *FIN* locus also displayed large effects on seed yield (19 and 15% of the phenotypic variance for SY trait in the MA and AM RIL populations, respectively). In the MA RIL population, the *FIN* locus also affected bracteole, leaf, pod, and seed size with a small to moderate additive effect (5 QTLs up to 3% of the phenotypic variance); whereas in AM RIL population, it affected pod size (9% of the phenotypic variance for PWI). In addition to the *FIN* locus, other genomic regions were also involved in the regulation of rate of plant production and time to flowering traits, although with a minor effect. Thus, for example, QTLs for PGT and PST in the MA population and for PT in the AM population were detected on Pv02 (PVESTBR006-GATS91), which colocalized with a QTL for seed weight previously mapped by Blair et al. (2006). Taken together, comparative QTL analysis results indicated that vegetative growth has a large effect on time to flowering and the rate of plant production traits, explaining the pleiotropic effects observed for FT and LMS traits. It is mainly due to the phenotypic effect of the recessive allele at the *FIN* locus, which is present in the determinate type I genotype (Beluga) and controls the meristem switch from a vegetative to a reproductive state. The *fin* allele reduces the plant growing period, causing a reduction in length of the main stem and number of branches, and an increase in internode length, as well as small bracteoles and leaves, large pods, and seeds that give rise to a lower yield (Norton, 1915; Ojehomon and Morgan, 1969; Koinange et al., 1996), resulting in common bean cultivars with a reduced flowering period since they mature more rapidly (Cober and Tanner, 1995; Koinange et al., 1996).

Genetic and Molecular Mechanisms Underlying the Link between Time to Flowering and Vegetative Growth

Currently, there are genes underlying flowering time QTLs which have been isolated in other crops; said crops have a growth pattern similar to common bean through successive series of modular units. In a common bean plant with a determinate type I growth habit, after floral initiation, the terminal shoot meristem produces a terminal inflorescence and ceases its vegetative growth. However, in a common bean plant exhibiting an indeterminate type IV growth habit, the terminal shoot meristem produces stem nodes, each one composed by one compound leaf and an inflorescence in its axil; thus, vegetative, and reproductive structures are continuously produced until maturity and senescence. This growth pattern through successive series of modular units is similar to that of

tomato (Sage and Webster, 1987; Schmitz and Theres, 1999). In tomato, *SELF-PRUNING* (*SP*) and *FALSIFLORA* (*FA*) control meristem identity. *SP* gene suppresses the transition of vegetative to reproductive state, keeping a plant indeterminate (Pnueli et al., 1998). *FA* is responsible for floral meristem identity and promotes flowering (Molinero-Rosales et al., 1999). *SP* is the homolog of the *Antirrhinum majus* *CENTRORADIALIS* (*CEN*) and *A. thaliana* *TERMINAL FLOWER1* (*TFL1*) genes (Pnueli et al., 2001). *FA* is an ortholog of *A. thaliana* *LFY* and *A. majus* *FLORICAULA* (*FLO*). *LFY* in *A. thaliana* activates directly *API*, causing flowering (Komeda, 2004; Saddic et al., 2006). Foucher et al. (2003) found that a pea homolog of the *A. thaliana* *TFL1*, *PsTFL1a*, corresponds to the determinacy locus (*DETERMINATE*; *DET*), and that another pea *TFL1* homolog, *PsTFL1c* (*LATE FLOWERING*; *LF*), acts as a repressor of flowering. Likewise, Mir et al. (2014) revealed the orthologous nature of *CcTFL1* gene for determinacy in pigeonpea; whereas in soybean, Tian et al. (2010) showed evidence that *Dt1* (*GmTFL1*) is a homolog of the *Arabidopsis* *TFL1* gene, which has a high-level of conservation with the common bean *PvTFL1y* gene. This gene has been proposed as a candidate gene for the *FIN* locus since mutations at the *PvTFL1y* locus were found to cosegregate with the determinate growth habit phenotype (Repinski et al., 2012). In this study, the position of *PvTFL1y* (*Phvul.001G189200*) was found within the QTLs positioned in the marker intervals of *FIN*-BMC224 and BMD045-*FIN* on Pv01. This finding is consistent with evidence at Andean *PvTFL1y* haplotype of the determinacy locus (Repinski et al., 2012). However, more information is needed to know whether different *PvTFL1y* haplotypes derived independently in each gene pool or whether the determinacy locus arose in a single gene pool, as happens in rice. In this species, the determinacy locus arose in a unique gene pool (*indica* or *japonica*) and was later transferred to the other pools (Sweeney et al., 2007). Furthermore, the correlated effects of the *PvTFL1y* locus on other plant traits, such as length of main stem or internode length, and productivity have been demonstrated in this work.

FUTURE REMARKS

The information herein reported could be used not only to establish different breeding strategies combining loci from the different gene pools of common bean, but also to look for associations of genetic variation in determinacy candidate genes in other legume crops with varieties bred for determinate growth habit, such as *P. coccineus* (runner bean), *P. lunatus* (lima bean), and *Cajanus cajan* (pigeonpea) (Waldia and Singh, 1987; van Rheenen et al., 1994; Huyghe, 1998). Furthermore, exploring the interaction and linkage of loci for vegetative growth, flowering time, and the rate of plant production may allow for the expansion of common bean to the geographic locations in which novel adaptation traits can be evaluated.

AUTHOR CONTRIBUTIONS

AG carried out molecular marker and QTL analysis and drafted the manuscript. FY supported mapping methodologies and

contributed to a critical review of the manuscript. SS performed the phenotypic data collection. SB contributed to molecular marker analysis. AD contributed to the review of the manuscript. RL collaborated in experimental design and critical review of the manuscript. MS planned the research work, assisted in analysis and interpretation of the data, and edited the manuscript. All authors have read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

This work was financially supported by the Ministerio de Economía y Competitividad (AGL2011-25562 and

AGL2014-51809 projects), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (RF2012-00026-C02-01 and RF2012-00026-C02-02 projects), Junta de Andalucía (Grant P12-AGR-01482 funded by Programa de Excelencia), and UE-FEDER Program. The authors would also like to thank Campus de Excelencia Internacional Agroalimentario-CeiA3 for partially supporting this work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01940/full#supplementary-material>

REFERENCES

- Aastveit, A. H., and Aastveit, K. (1993). Effects of genotype-environment interactions on genetic correlations. *Theor. Appl. Genet.* 86, 1007–1013. doi: 10.1007/BF00211054
- Abe, M., Kobayashi, Y., Yamamoto, S., Daimon, Y., Yamaguchi, A., Ikeda, Y., et al. (2005). FD, a bZIP protein mediating signals from the floral pathway integrator FT at the shoot apex. *Science* 309, 1052–1056. doi: 10.1126/science.1115983
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Bassett, M. J. (1997). Tight linkage between the Fin locus for plant habit and the Z locus for partly colored seed coat patterns in common bean. *J. Am. Soc. Hortic. Sci.* 122, 656–658.
- Benchimol, L. L., Campos, T., Carbonell, S. A. M., Colombo, C. A., Chioratto, A. F., Formighieri, E. F., et al. (2007). Structure of genetic diversity among common bean (*Phaseolus vulgaris* L.) varieties of Mesoamerican and Andean origins using new developed microsatellite markers. *Genet. Resour. Crop. Evol.* 54, 1747–1762. doi: 10.1007/s10722-006-9184-3
- Benlloch, R., Berbel, A., Ali, L., Gohari, G., Millán, T., and Madueño, F. (2015). Genetic control of inflorescence architecture in legumes. *Front. Plant Sci.* 6:543. doi: 10.3389/fpls.2015.00543
- Blair, J. E., Coffey, M. D., Park, S. Y., Geiser, D. M., and Kang, S. (2008). A multi-locus phylogeny for Phytophthora utilizing markers derived from complete genome sequences. *Fungal Genet. Biol.* 45, 266–277. doi: 10.1016/j.fgb.2007.10.010
- Blair, M. W., Hurtado, N., Chavarro, C. M., Muñoz-Torres, M., Giraldo, M. C., Pedraza, F., et al. (2011). Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMC series. *BMC Plant Biol.* 11:50. doi: 10.1186/1471-2229-11-50
- Blair, M. W., Iriarte, G., and Beebe, S. (2006). QTL analysis of yield traits in an advanced backcross population derived from a cultivated Andean × wild common bean (*Phaseolus vulgaris* L.) cross. *Theor. Appl. Genet.* 112, 1149–1163. doi: 10.1007/s00122-006-0217-2
- Blair, M. W., Medina, J., Astudillo, C., Rengifo, J., Beebe, S., Machado, G., et al. (2010). QTL for seed iron and zinc concentrations in a recombinant inbred line population of Mesoamerican common beans (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 121, 1059–1071. doi: 10.1007/s00122-010-1371-0
- Blair, M. W., Mu-oz-Torres, M., Giraldo, M. C., and Pedraza, F. (2009b). Development and diversity assessment of Andean-derived, gene-based microsatellites for common bean (*Phaseolus vulgaris* L.). *BMC Plant Biol.* 9:100. doi: 10.1186/1471-2229-9-100
- Blair, M. W., Pedraza, F., Giraldo, M. C., Buendía, H. F., and Hurtado, N. (2009a). Development of microsatellite markers for common bean (*Phaseolus vulgaris* L.) based on screening of non-enriched small insert genomic libraries. *Genome* 52, 772–787. doi: 10.1139/G09-053
- Blair, M. W., Pedraza, F., Buendía, H. F., Gaitán-Solis, E., Beebe, S. E., Gepts, P., et al. (2003). Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 107, 1362–1374. doi: 10.1007/s00122-003-1398-6
- Borel, J. C., Ramalho, M. A., and Abreu, A. F. (2016). Epistasis in intra- and inter-gene pool crosses of the common bean. *Genet. Mol. Res.* 15:15017573. doi: 10.4238/gmr.15017573
- Boss, P. K., Bastow, R. M., Mylne, J. S., and Dean, C. (2004). Multiple pathways in the decision to flower: enabling, promoting, and resetting. *Plant Cell* 16, S18–S31. doi: 10.1105/tpc.015958
- Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R., and Coen, E. (1997). Inflorescence commitment and architecture in Arabidopsis. *Science* 275, 80–83. doi: 10.1126/science.275.5296.80
- Broughton, W. J., Hernandez, G., Blair, M., Beebe, S., Gepts, P., and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) - model food legumes. *Plant Soil* 252, 55–128. doi: 10.1023/A:1024146710611
- Bruzzi, A. T., Ramalho, M. A. P., and Ferreira, D. F. (2007). Homeostasis of common bean populations with different genetic structures. *Crop Breed. Appl. Biotechnol.* 2, 111–116. doi: 10.12702/1984-7033.v07n02a01
- Buso, G. S. C., Amaral, Z. P. S., Brondani, R. P. V., and Ferreira, M. E. (2006). Microsatellite markers for the common bean *Phaseolus vulgaris*. *Mol. Ecol. Notes* 6, 252–254. doi: 10.1111/j.1471-8286.2006.01210.x
- Campos, T., Oblessuc, P. R., Sforça, D. A., Cardoso, J. M. K., Baroni, R. M., Sousa, A. C. B., et al. (2011). Inheritance of growth habit detected by genetic linkage analysis using microsatellites in the common bean (*Phaseolus vulgaris* L.). *Mol. Breeding* 27, 549–560. doi: 10.1007/s11032-010-9453-x
- Cardoso, J. M. K., Oblessuc, P. R., Campos, T., Sforça, D. A., Carbonell, S. A. M., Chioratto, A. F., et al. (2008). New microsatellite markers developed from an enriched microsatellite common bean library. *Pesq. Agropec. Bras.* 43, 929–936. doi: 10.1590/S0100-204X2008000700019
- Chavarro, M. C., and Blair, M. W. (2010). QTL Analysis and effect of the fin locus on tropical adaptation in an inter-genepool common bean population. *Trop. Plant Biol.* 3, 204–218. doi: 10.1007/s12042-010-9058-x
- Checa, O., and Blair, M. W. (2008). Mapping QTL for climbing ability and component traits in common bean (*Phaseolus vulgaris* L.). *Mol. Breeding* 22, 201–215. doi: 10.1007/s11032-008-9167-5
- Checa, O., Ceballos, H., and Blair, M. W. (2006). Generation means analysis of climbing ability in common bean (*Phaseolus vulgaris* L.). *J. Hered.* 97, 456–465. doi: 10.1093/jhered/esl025
- Chen, D. H., and Ronald, P. C. (1999). A rapid DNA miniprep method suitable for AFLP and other PCR applications. *Plant Mol. Biol. Rep.* 17, 53–57. doi: 10.1023/A:1007585532036
- Cichy, K. A., Blair, M. W., Galeno, C., Snapp, S. S., and Kelly, J. D. (2009). QTL analysis of root architecture traits and low phosphorus tolerance in an Andean bean population. *Crop Sci.* 49, 59–68. doi: 10.2135/cropsci2008.03.0142
- Cober, E. R., and Tanner, J. W. (1995). Performance of related indeterminate and tall determinate soybean lines in short-season areas. *Crop Sci.* 35, 361–364. doi: 10.2135/cropsci1995.0011183X003500020011x
- Córdoba, J. M., Chavarro, C., Schlueter, J. A., Jackson, S. A., and Blair, M. W. (2010). Integration of physical and genetic maps of common

- bean through BAC-derived microsatellite markers. *BMC Genomics* 11:436. doi: 10.1186/1471-2164-11-436
- Davis, D. W., and Frazier, W. A. (1966). Inheritance of some growth habit components in certain types of bush lines of *Phaseolus vulgaris* L. *Proc. Amer. Soc. Hort. Sci.* 88, 384–392.
- Debouck, D. G. (1991). “Systematics and morphology,” in *Common Beans: Research for Crop Improvement*, eds A. van Schoonhoven and O. Voysest (Wallingford: Commonwealth Agricultural Bureaux International), 55–118.
- Debouck, D. G., and Hidalgo, R. (1986). *Morphology of the Common Bean Plant Phaseolus vulgaris*. Cali: Centro Internacional de Agricultura Tropical.
- Dhanasekar, P., and Reddy, K. S. (2014). A novel mutation in TFL1 homolog affecting determinacy in cowpea (*Vigna unguiculata*). *Mol. Genet. Genomics* 290, 55–65. doi: 10.1007/s00438-014-0899-0
- Foucher, F., Morin, J., Courtiade, J., Cadioux, S., Ellis, N., Bandfield, M. J., et al. (2003). DETERMINATE and LATE FLOWERING are two TERMINAL FLOWER1/CENTRORADIALIS homologs that control two distinct phases of flowering initiation and development in pea. *Plant Cell* 15, 2742–2754. doi: 10.1105/tpc.015701
- Frazier, W. A., Baggett, J. R., and Sistrunk, W. A. (1958). Transfer of certain Blue Lake bean pod characters to bush beans. *Proc. Am. Soc. Hort. Sci.* 71, 416–421.
- Freyer, R., Skroch, P., Geffroy, V., Adam-Blondon, A. F., Shirmohamadi, A., Johnson, W. C., et al. (1998). Towards an integrated linkage map of common bean. 4. Development of a core map and alignment of RFLP maps. *Theor. Appl. Genet.* 97, 847–856. doi: 10.1007/s001220050964
- Gaitán-Solis, E., Duque, M. C., Edwards, K. J., and Tohme, J. (2002). Microsatellite repeats in common bean (*Phaseolus vulgaris*) isolation, characterization, and cross-species amplification in *Phaseolus* spp. *Crop Sci.* 42, 2128–2136. doi: 10.2135/cropsci2002.2128
- García, R. A., Rangel, P. N., Brondani, C., Martins, W. S., Melo, L. C., Carneiro, M. S., et al. (2011). The characterization of a new set of EST-derived simple sequence repeat (SSR) markers as a resource for the genetic analysis of *Phaseolus vulgaris*. *BMC Genet.* 12:41. doi: 10.1186/1471-2156-12-41
- Gepts, P., and Debouck, D. (1991). “Origin, domestication, and evolution of the common bean (*Phaseolus vulgaris* L.),” in *Common Beans: Research for Crop Improvement*, ed A. van Schoonhoven (Cali: CIAT), 7–53.
- Grisi, M. C. M., Blair, M. W., Gepts, P., Brondani, C., Pereira, P. A. A., and Brondani, R. P. V. (2007). Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT93 × Jalo EEP558. *Genet. Mol. Res.* 6, 691–706.
- Hanai, L. R., de Campos, T., Camargo, L. E., Benchimol, L. L., Souza, A. P., Melotto, M., et al. (2007). Development, characterization, and comparative analysis of polymorphism at common bean SSR loci isolated from genic and genomic sources. *Genome* 50, 266–277. doi: 10.1139/G07-007
- Holland, J. B. (2006). Estimating genotypic correlations and their standard errors using multivariate restricted maximum likelihood estimation in SAS Proc MIXED. *Crop Sci.* 46, 642–654. doi: 10.2135/cropsci2005.0191
- Holland, J. B., Nyquist, W. E., and Cervantes-Martínez, C. T. (2003). Estimating and interpreting heritability for plant breeding: an update. *Plant Breeding Rev.* 22, 9–112. doi: 10.1002/9780470650202.ch2
- Hougaard, B. K., Madsen, L. H., Sandal, N., de Carvalho Moretzsohn, M., Fredslund, J., Schause, L., et al. (2008). Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and *Arachis*. *Genetics* 179, 2299–2312. doi: 10.1534/genetics.108.090084
- Huyghe, C. (1998). Genetics and genetic modifications of plant architecture in grain legumes: a review. *Agronomie* 18, 383–411. doi: 10.1051/agro:19980505
- Johnson, W. C., and Gepts, P. (1999). Segregation for performance in recombinant inbred populations resulting from inter-gene pool crosses of common bean (*Phaseolus vulgaris* L.). *Euphytica* 106, 45–56. doi: 10.1023/A:1003541201923
- Johnson, W. C., and Gepts, P. (2002). The role of epistasis in controlling seed yield and other agronomic traits in an Andean × Mesoamerican cross of common bean (*Phaseolus vulgaris* L.). *Euphytica* 125, 69–79. doi: 10.1023/A:1015775822132
- Jung, G., Coyne, D., Skroch, P., Nienhuis, J., Arnaud-Santana, E., Bokosi, J., et al. (1996). Molecular markers associated with plant architecture and resistance to common blight, web blight, and rust in common beans. *J. Am. Soc. Hort. Sci.* 121, 794–803.
- Kelly, J. D., Hosfield, G. L., Varner, G. V., Uebersax, M. A., and Taylor, J. (1999). Registration of ‘Beluga’ Alubia Bean. *Crop Sci.* 39, 294. doi: 10.2135/cropsci1999.0011183X003900010065x
- Koinange, E., Singh, S., and Gepts, P. (1996). Genetic Control of the domestication syndrome in common bean. *Crop Sci.* 36, 1037–1045. doi: 10.2135/cropsci1996.0011183X003600040037x
- Kolkman, J. M., and Kelly, J. D. (2003). QTL conferring resistance and avoidance to white mold in common bean. *Crop Sci.* 43, 539–548. doi: 10.2135/cropsci2003.0539
- Komeda, Y. (2004). Genetic regulation of time to flower in *Arabidopsis thaliana*. *Annu. Rev. Plant Biol.* 55, 521–535. doi: 10.1146/annurev.arplant.55.031903.141644
- Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., et al. (2010). Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* 154, 1220–1231. doi: 10.1104/pp.110.160796
- Kornegay, J., White, J. W., and Ortiz de la Cruz, O. (1992). Growth habit and gene pool effects on inheritance of yield in common bean. *Euphytica* 62, 171–180. doi: 10.1007/BF00041751
- Kosambi, D. D. (1944). The estimation of map distances from recombination values. *Ann. Eugenics* 12, 172–175. doi: 10.1111/j.1469-1809.1943.tb02321.x
- Kwak, M., Toro, O., Debouck, D. G., and Gepts, P. (2012). Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*). *Ann. Bot.* 110, 1573–1580. doi: 10.1093/aob/mcs207
- Ma, H. (1998). To be, or not to be, a flower: control of floral meristem identity. *Trends Genet.* 14, 26–32. doi: 10.1016/S0168-9525(97)01309-7
- McClean, P. E., Lee, R., Otto, C., Gepts, P., and Bassett, M. J. (2002). Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *J. Hered.* 93, 148–152. doi: 10.1093/jhered/93.2.148
- McConnell, M., Mamidi, S., Lee, R., Chikara, S., Rossi, M., Papa, R., et al. (2010). Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 121, 1103–1116. doi: 10.1007/s00122-010-1375-9
- Mir, R. R., Kudapa, H., Srikanth, S., Saxena, R. K., Sharma, A., Azam, S., et al. (2014). Candidate gene analysis for determinacy in pigeonpea (*Cajanus* spp.). *Theor. Appl. Genet.* 127, 2663–2678. doi: 10.1007/s00122-014-2406-8
- Molinero-Rosales, N., Jamilena, M., Zurita, S., Gómez, P., Capel, J., and Lozano, R. (1999). FALSIFLORA, the tomato orthologue of FLORICAULA and LEAFY, controls flowering time and floral meristem identity. *Plant J.* 20, 685–693. doi: 10.1046/j.1365-3113X.1999.00641.x
- Montgomery, J., Wittwer, C. T., Palais, R., and Zhou, L. (2007). Simultaneous mutation scanning and genotyping by high-resolution DNA melting analysis. *Nature* 2, 59–66. doi: 10.1038/nprot.2007.10
- Moreto, A. L., Ramalho, M. A. P., and Bruzi, A. T. (2012). Epistasis in an Andean × Mesoamerican cross of common bean. *Euphytica* 186, 755–760. doi: 10.1007/s10681-011-0578-8
- Nilsson, O., Lee, I., Blázquez, M. A., and Weigel, D. (1998). Flowering-time genes modulate the response to LEAFY activity. *Genetics* 150, 403–410.
- Norton, J. B. (1915). Inheritance of habit in the common bean. *Am. Nat.* 49, 547–561. doi: 10.1086/279499
- Ohshima, S., Murata, M., Sakamoto, W., Ogura, Y., and Motoyoshi, F. (1997). Cloning and molecular analysis of the Arabidopsis gene TERMINAL FLOWER 1. *Mol. Gen. Genet.* 254, 186–194. doi: 10.1007/s004380050407
- Ojehomon, O. O., and Morgan, D. G. (1969). A quantitative study of inflorescence development in *Phaseolus vulgaris*. *Ann. Bot.* 33, 325–332.
- Pedrosa-Harand, A., Porch, T., and Gepts, P. (2008). Standard nomenclature for common bean chromosomes and linkage groups. *Annu. Rep. Bean. Improv. Coop.* 51, 106–107.
- Pérez-Vega, E., Pañeda, A., Rodríguez-Suárez, C., Campa, A., Giraldez, R., and Ferreira, J. (2010). Mapping of QTLs for morpho-agronomic and seed quality traits in a RIL population of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* 120, 1367–1380. doi: 10.1007/s00122-010-1261-5
- Pidkowich, M. S., Klenz, J. E., and Haughn, G. W. (1999). The making of a flower: control of floral meristem identity in Arabidopsis. *Trends Plant Sci.* 4, 64–70. doi: 10.1016/S1360-1385(98)01369-7
- Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganai, M., et al. (1998). The SELF-PRUNING gene of tomato regulates vegetative to

- reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1. *Development* 125, 1979–1989.
- Pnueli, L., Gutfinger, T., Hareven, D., Ben-Naim, O., Ron, N., Adir, N., et al. (2001). Tomato SP-interacting proteins define a conserved signaling system that regulates shoot architecture and flowering. *Plant Cell* 13, 2687–2702. doi: 10.1105/tpc.010293
- Repinski, S. L., Kwak, M., and Gepts, P. (2012). The common bean growth habit gene PvTFL1y is a functional homolog of Arabidopsis TFL1. *Theor. Appl. Genet.* 124, 1539–1547. doi: 10.1007/s00122-012-1808-8
- Saddic, L. A., Huvermann, B., Bezghani, S., Su, Y., Winter, C. M., Kwon, C. S., et al. (2006). The LEAFY target LMI1 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development* 133, 1673–1682. doi: 10.1242/dev.02331
- Sage, T. L., and Webster, B. D. (1987). Flowering and fruiting patterns of *Phaseolus vulgaris* L. *Bot. Gaz.* 148, 35–41. doi: 10.1086/337625
- Santalla, M., Monteagudo, A. B., González, A. M., Lema, M., De la Fuente, M., and De Ron, A. M. (2004). Agronomic potential value of great northern recombinant lines and breeding implications in common bean. *Annu. Rep. Bean. Improv. Coop.* 47, 147–148.
- Schmitz, G., and Theres, K. (1999). Genetic control of branching in Arabidopsis and tomato. *Curr. Opin. Plant Biol.* 2, 51–55. doi: 10.1016/S1369-5266(99)80010-7
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G., Cannon, S., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Singh, S. P., and Urrea, C. A. (1995). Inter- and intraracial hybridization and selection for seed yield in early generations of common bean, *Phaseolus vulgaris* L. *Euphytica* 81, 31–137. doi: 10.1007/BF00025424
- Sweeney, M. T., Thomson, M. J., Cho, Y. G., Park, Y. J., Williamson, S. H., Bustamante, C. D., et al. (2007). Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet.* 3, 1418–1424. doi: 10.1371/journal.pgen.0030133
- Tanaka, A., and Fujita, K. (1979). Growth, photosynthesis and yield components in relation to grain yield of the field bean. *J. Faculty Agric. Hokkaido Univ.* 59, 145–238.
- Tar'an, B., Michaels, T. E., and Pauls, K. P. (2002). Genetic mapping of agronomic traits in common bean. *Crop Sci.* 42, 544–556. doi: 10.2135/cropsci2002.0544
- Tian, Z., Wang, X., Lee, R., Li, Y. H., Specht, J. E., Nelson, R. L., et al. (2010). Artificial selection for determinate growth habit in soybean. *Proc. Natl. Acad. Sci. U.S.A.* 107, 8563–8568. doi: 10.1073/pnas.1000088107
- van Ooijen, J. W. (2006). *JoinMap® 4, Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma B.V.
- van Rheenen, H. A., Pundir, R. P. S., and Miranda, J. H. (1994). Induction and inheritance of determinate growth habit in chickpea (*Cicer arietinum* L.). *Euphytica* 78, 137–141. doi: 10.1007/BF00021409
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Waldia, R. S., and Singh, V. P. (1987). Inheritance of stem termination in Pigeon Pea (*Cajanus cajan* (L.) Mill sp.). *Euphytica* 36, 525–527. doi: 10.1007/BF00041497
- Wallace, D., Yourstone, K., Masaya, P., and Zobel, R. (1993). Photoperiod gene control over partitioning between reproductive vs. vegetative growth. *Theor. Appl. Genet.* 86, 6–16. doi: 10.1007/BF00223803
- Wang, C. S., Rutledge, J. J., and Gianola, D. (1994). Bayesian analysis of mixed linear models via Gibbs sampling with an application to litter size in Iberian pigs. *Genet. Sel. Evol.* 26, 91–115. doi: 10.1186/1297-9686-26-2-91
- Weigel, D. (1995). The genetics of flower development: from floral induction to ovule morphogenesis. *Annu. Rev. Plant Phys.* 29, 19–39. doi: 10.1146/annurev.ge.29.120195.000315
- Weller, J. L., and Ortega, R. (2015). Genetic control of flowering time in legumes. *Front. Plant Sci.* 6:207. doi: 10.3389/fpls.2015.00207
- Wigge, P. A., Kim, M. C., Jaeger, K. E., Busch, W., Schmid, M., Lohmann, J. U., et al. (2005). Integration of spatial and temporal information during floral induction in Arabidopsis. *Science* 309, 1056–1059. doi: 10.1126/science.1114358
- Wright, E., and Kelly, J. (2011). Mapping QTL for seed yield and canning quality following processing of black bean (*Phaseolus vulgaris* L.). *Euphytica* 179, 471–484. doi: 10.1007/s10681-011-0369-2
- Yang, J., Hu, C., Hu, H., Yu, R., Xia, Z., Ye, X., et al. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24, 721–723. doi: 10.1093/bioinformatics/btm494
- Yanofsky, M. F. (1995). Floral meristems to floral organs: genes controlling early events in Arabidopsis flower development. *Annu. Rev. Plant Phys.* 46, 167–188. doi: 10.1146/annurev.pp.46.060195.001123
- Yu, K., Park, S. J., Poysa, V., and Gepts, P. (2000). Integration of simple sequence repeat (SSR) markers into a molecular linkage map of common bean (*Phaseolus vulgaris* L.). *J. Hered.* 91, 429–434. doi: 10.1093/jhered/91.6.429

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 González, Yuste-Lisbona, Saburido, Bretones, De Ron, Lozano and Santalla. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Gene Mapping of a Mutant Mungbean (*Vigna radiata* L.) Using New Molecular Markers Suggests a Gene Encoding a YUC4-like Protein Regulates the Chasmogamous Flower Trait

Jingbin Chen^{1,2†}, Prakrit Somta^{2†}, Xin Chen^{1*}, Xiaoyan Cui¹, Xingxing Yuan¹ and Peerasak Srinives^{2*}

OPEN ACCESS

Edited by:

Nicolas Rispail,
Institute for Sustainable
Agriculture-CSIC, Spain

Reviewed by:

Vikas Kumar Singh,
International Crops Research Institute
for the Semi Arid Tropics, India
Rupesh Kailasrao Deshmukh,
Laval University, Canada

*Correspondence:

Peerasak Srinives
agrps@yahoo.com;
Xin Chen
cx@jaas.ac.cn

[†] These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 12 March 2016

Accepted: 26 May 2016

Published: 10 June 2016

Citation:

Chen J, Somta P, Chen X, Cui X,
Yuan X and Srinives P (2016) Gene
Mapping of a Mutant Mungbean
(*Vigna radiata* L.) Using New
Molecular Markers Suggests a Gene
Encoding a YUC4-like Protein
Regulates the Chasmogamous
Flower Trait. *Front. Plant Sci.* 7:830.
doi: 10.3389/fpls.2016.00830

¹ Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences, Nanjing, China, ² Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen, Kasetsart University, Nakhon Pathom, Thailand

Mungbean (*Vigna radiata* L.) is a cleistogamous plant in which flowers are pollinated before they open, which prevents yield improvements through heterosis. We previously generated a chasmogamous mutant (CM) mungbean in which open flowers are pollinated. In this study, we developed insertion/deletion (indel) markers based on the transcriptome differences between CM and Sulu-1 (i.e., normal flowering) plants. An F₂ population derived from a cross between CM and Sulu-1 was used for gene mapping. Segregation analyses revealed that a single recessive gene regulates the production of chasmogamous flowers. Using newly developed indel and simple sequence repeat markers, the *cha* gene responsible for the chasmogamous flower trait was mapped to a 277.1-kb segment on chromosome 6. Twelve candidate genes were detected in this segment, including *Vradi06g12650*, which encodes a YUCCA family protein associated with floral development. A single base pair deletion producing a frame-shift mutation and a premature stop codon in *Vradi06g12650* was detected only in CM plants. This suggested that *Vradi06g12650* is a *cha* candidate gene. Our results provide important information for the molecular breeding of chasmogamous mungbean lines, which may serve as new genetic resources for hybrid cultivar development.

Keywords: mungbean, *Vigna radiata*, chasmogamy, outcrossing, gene mapping, hybrid development

INTRODUCTION

Exploiting heterosis is an effective way to increase crop yield. As an important legume in Asia, several studies have been conducted on mungbean (*Vigna radiata* L. Wilczek) to investigate the importance of heterosis for seed yield and other yield-related traits (Chen et al., 2003; Soehendi and Srinives, 2005; Sorajjapinun et al., 2012; Yashpal et al., 2015). However, a major obstacle to producing mungbean hybrid seeds is floral architecture. Mungbean plants have papilionaceous flowers with five differently shaped petals, including one standard petal, two wing petals, and two keel petals (Verdcourt, 1970). The anthers and stigmas are enclosed within the pocket-like keel

petals, making mungbean a cleistogamous plant whose flowers are pollinated before they open. The natural outcrossing rate of cultivated mungbean is approximately 1.68% (Sangiri et al., 2007), which is too low for the commercial production of hybrid seeds. Some accessions of cleistogamous plants exhibit a specific floral architecture that promotes natural outcrossing. Studies in rice revealed that exerted stigmas and anthers increase the outcrossing rate and considerably help the production of hybrid seeds (Kato and Namai, 1987). Some mutant legume crops with exerted stigmas and anthers have been described (Dahiya et al., 1984; Pundir and Reddy, 1998; Cherian et al., 2006; Srinivasan and Gaur, 2012; Dheer et al., 2014). We previously identified a chasmogamous mutant (CM), which was generated in the mungbean accession V1197 by gamma irradiation (Sorajjapinun and Srinives, 2011). The outcrossing rate of CM plants increased to 9.6%, while outcrossing was undetectable in wild-type controls. Additionally, the yield and agronomic traits of CM plants differed from those of V1197 plants, with fewer pods per plant, seeds per pod, and yield per plant. Genetic analyses using F₁, F₂, and backcross populations from the cross between CM and V1197 plants revealed that the production of chasmogamous flowers is regulated by a single gene, *cha* (Sorajjapinun and Srinives, 2011). Thus, additional research focusing on the genetic characterization of this mutant mungbean should be conducted.

In terms of molecular genetics and genomics, mungbean has not been as extensively studied as other legumes, including soybean [*Glycine max* (L.) Merr.], common bean (*Phaseolus vulgaris* L.), cowpea [*V. unguiculata* (L.) Walp.], and chickpea (*Cicer arietinum* L.). DNA markers represent important tools for genetic analyses and the mapping of genes or quantitative trait loci. Many DNA markers, especially simple sequence repeats (SSRs), have been developed for mungbean (Gwag et al., 2007; Somta et al., 2008, 2009; Tangphatsornruang et al., 2009; Gupta et al., 2014; Chen et al., 2015) or introduced from other closely related species. These markers have been used to develop linkage maps and map quantitative trait loci associated with important mungbean traits (e.g., yield and resistance to biotic and abiotic stresses; Kasettranant et al., 2010; Chankaew et al., 2011, 2013; Isemura et al., 2012; Kajonphol et al., 2012; Prathet et al., 2012; Sompong et al., 2012; Kitsanachandee et al., 2013; Alam A.K.M.M. et al., 2014; Alam A.M. et al., 2014; Chotechung et al., 2016). However, most of these markers are monomorphic or weakly polymorphic. One of the most important advances in mungbean genomics research was the release of the VC1973A whole genome sequence by Kang et al. (2014). This sequence is relevant for research related to marker development, gene mapping, and gene function analyses.

The molecular mechanisms regulating floral development can be summarized using an “ABCE” model (Krizek and Fletcher, 2005). Additionally, some hormones (e.g., auxin) are important for floral development (Okada et al., 1991; Bennett et al., 1995; Przemeck et al., 1996; Galweiler et al., 1998; Aloni et al., 2006; Cheng et al., 2006). Some *Arabidopsis thaliana* mutants have abnormal flowers, and the genes responsible for this mutation have been identified as *pin1*, *pinoid*, *mp*, and *yuc* (Cheng and Zhao, 2007). The YUCCA (YUC) family consists of flavin monooxygenases (FMOs) related to the

biosynthesis of indole-3-acetic acid. The FMOs are key enzymes that convert indole-3-pyruvic acid to indole-3-acetic acid by catalyzing the hydroxylation of the amino group of tryptophan, which is the rate-limiting step in tryptophan-dependent auxin biosynthesis (Zhao et al., 2001). There are 11 *YUC* genes in the *A. thaliana* genome (Zhao et al., 2001; Cheng et al., 2006, 2007). Double, triple, and quadruple mutants of some *YUC* family genes exhibit severe defects in floral patterns, vascular formations, and other developmental processes (Cheng et al., 2006). Therefore, *YUC* genes play important roles during floral formation and development, with implications for flower shape.

In this study, we mapped the *cha* gene regulating the production of chasmogamous flowers in CM plants using new insertion/deletion (indel) and SSR markers. We identified a gene encoding a YUC-like protein as a likely candidate for the *cha* gene.

MATERIALS AND METHODS

Mapping Population and DNA Extraction

We previously identified a CM mungbean line in the M₂ generation of accession V1197 following gamma irradiation. The CM plants lacked wing and keel petals, which exposed the stigmas and anthers (Sorajjapinun and Srinives, 2011). A stable CM line was selected from the M₃ and M₄ generations. A CM plant was pollinated using pollen from Sulu-1, which is a mungbean line with normal flowers. Three F₁ plants were grown and the resulting flowers were morphologically analyzed. The hybrids were verified using two polymorphic indel markers (i.e., VRID001 and VRID002; Supplementary Table S1). Seeds from one F₁ plant were then harvested to produce an F₂ population consisting of 127 plants. The F₂ plants and their parents were grown in a field at Kasetsart University, Kamphaeng Saen Campus, Nakhon Pathom, Thailand from May to July 2014. Flowers from each plant were individually examined to determine whether they were normal or chasmogamous.

Total genomic DNA was extracted from fresh leaf tissue of individual plants according to a slightly modified version of the method described by Lodhi et al. (1994). All DNA samples were diluted to 5 ng μl^{-1} according to lambda DNA, and analyzed by 1.0% agarose gel electrophoresis.

Transcriptome Sequencing and Development of Molecular Markers

Total RNA was extracted from young flowers of CM and Sulu-1 mungbean plants using the EasyPure Plant RNA kit (Transgene Biotech, Beijing, China). The RNA samples were used to prepare libraries for sequencing by the Illumina HiSeq 2000 sequencer at the Beijing Genomics Institute (Shenzhen, China). The resulting sequences were assembled using the Trinity program (Grabherr et al., 2011). The Sulu-1 and CM transcriptome sequences were aligned using the NCBI BLAST+ 2.2.31 program with an *E*-value cutoff of 10.0. Sequences with indels that were 5 bp or larger were randomly chosen for marker development. Primers

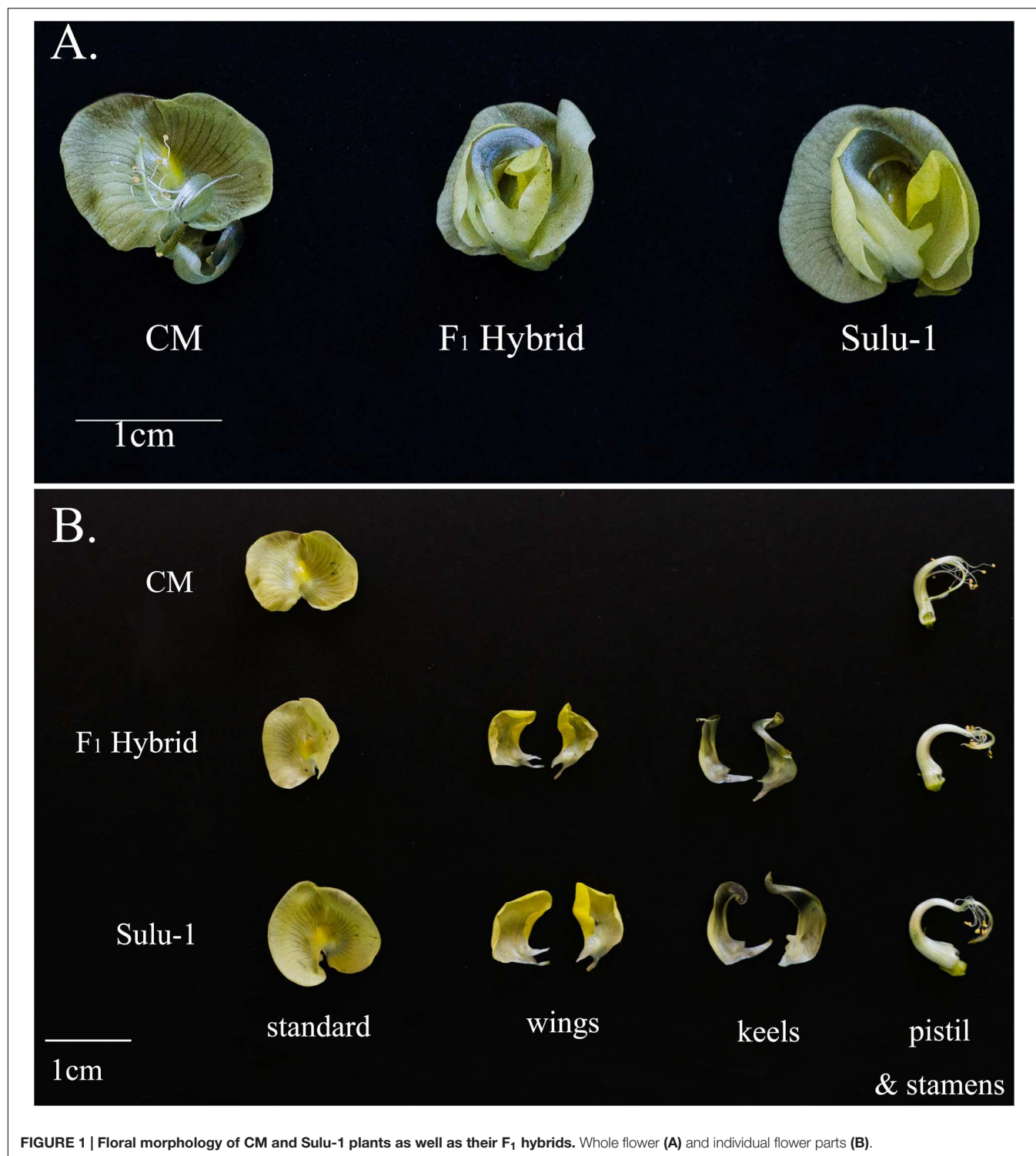


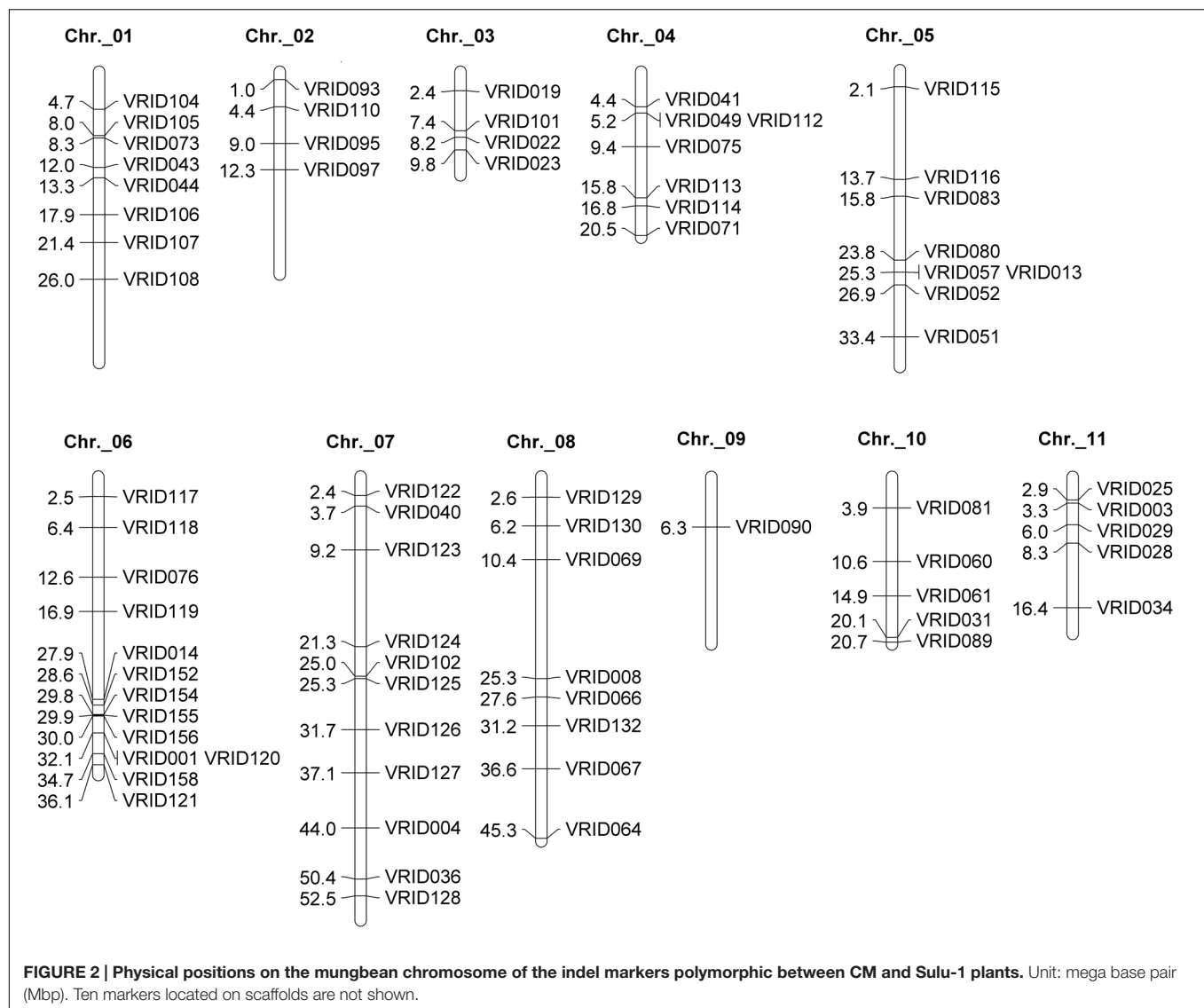
FIGURE 1 | Floral morphology of CM and Sulu-1 plants as well as their F₁ hybrids. Whole flower (A) and individual flower parts (B).

specific for the selected indels were designed using Primer3 (Untergasser et al., 2012) with the following criteria: primer length: 18–27 nucleotides; melting temperature: 50–65°C; GC content: 40–60%; and polymerase chain reaction (PCR) product size: 100–300 bp. Transcript sequences used for indel marker development were blasted against the mungbean whole genome

sequence¹ (Kang et al., 2014) to determine the physical locations of markers.

In addition to indel markers, we developed new SSR markers to fine map the *cha* locus. The mungbean chromosome 6

¹http://plantgenomics.snu.ac.kr/mediawiki-1.21.3/index.php/Main_Page



sequence was downloaded (Kang et al., 2014) and scanned for di-, tri-, and tetra-nucleotide repeats using SSR Hunter 1.3 (Li and Wan, 2005). Based on our initial mapping of the *cha* locus using indel markers, we focused on a 2.2-Mb genomic region of chromosome 6 (i.e., 29.9–32.1 Mb) carrying *cha*, and selected the SSRs therein. Primers for the SSRs were designed as described for the indel markers.

Molecular Marker Analysis

The newly developed indel and SSR markers were used to detect polymorphisms between CM and Sulu-1 sequences. The PCR analyses were completed using a 10-μl solution containing 2 ng genomic DNA, 1x *Taq* buffer, 2 mM MgCl₂, 0.2 mM dNTPs, 1 U *Taq* DNA polymerase (Thermo Scientific), and 0.5 μM forward and reverse primers. The PCR was conducted in an MJ Research PTC-200 Thermal Cycler (Bio-Rad) using the following program: 94°C for 3 min; 35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30s; 72°C for 5 min. The amplicons were

separated in a 6% denaturing polyacrylamide gel or a 3% agarose gel, and visualized with silver or ethidium bromide staining, respectively. The polymorphic markers were used to analyze the F₂ population.

Data Analysis and Gene Mapping

We counted the number of plants with normal and chasmogamous flowers, and analyzed the data using a χ^2 test (Mather, 1951) to confirm the monogenic inheritance in CM plants as reported earlier by Sorajjapinun and Srinives (2011). Segregation data for the floral traits and DNA markers were used to construct a linkage map with MapMaker/EXP 3.0 (Lander et al., 1987). A minimum logarithm of odds value of 3.0 and maximum recombination frequency of 4.0 were used to group the markers. The genetic map distance was calculated using the Kosambi mapping function. Linkage groups were anchored to chromosomes by the physical location of markers. The map was drawn using MapChart 2.30 (Voorrips, 2002).

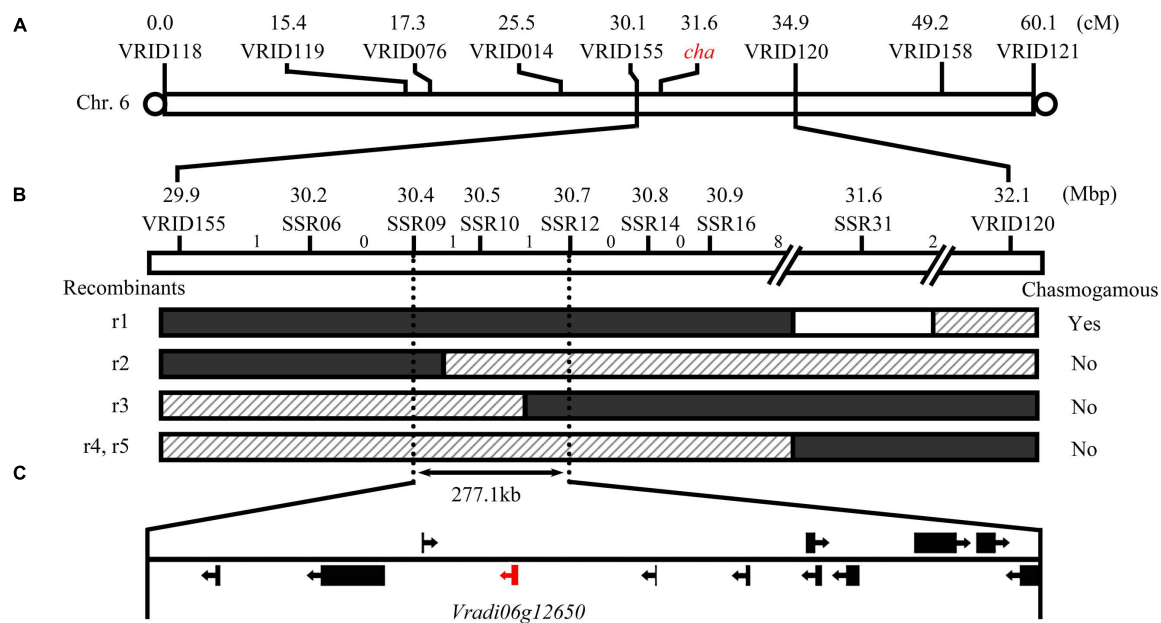


FIGURE 3 | Mapping of the *cha* gene regulating the production of chasmogamous flowers in CM plants. Location of *cha* (A) and recombinant events on chromosome 6 (B). White, shadow, and black regions in the recombinants indicate homozygous regions for the Sulu-1 allele, and heterozygous and homozygous regions for the CM allele, respectively. The number of recombinant events between adjacent markers are indicated above the bar. Another five recombinants between the homozygous Sulu-1 genotype and heterozygous genotype are not presented because all of the plants had normal flowers. The 12 predicted genes in the candidate gene region are indicated (C).

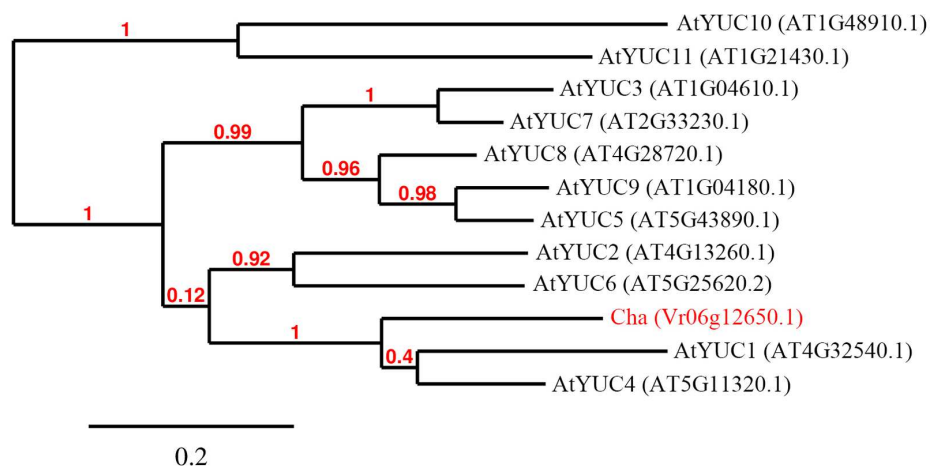
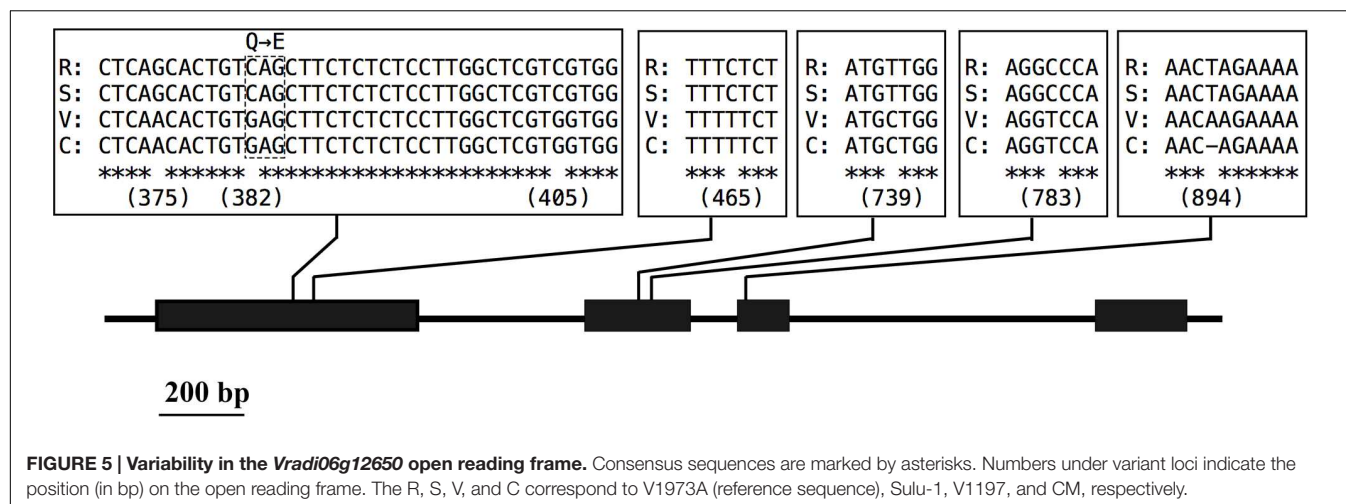


FIGURE 4 | Phylogenetic tree of *Arabidopsis thaliana* YUCCA and *Cha* proteins. Sequences of 11 *A. thaliana* YUCCA proteins and the deduced mungbean Vr06g12650.1 protein were used to generate the phylogenetic tree.

Identifying and Sequencing the Candidate Gene

Based on the locations of the markers flanking *cha* on the linkage map, the predicted genes on the mungbean reference sequence (Kang et al., 2014) between the flanking markers were downloaded. Deduced protein sequences for these genes were subjected to a BLASTP search against the NCBI database to obtain information regarding their homologs and functions.

The following two pairs of primers were designed to amplify the genomic region containing a candidate gene responsible for the production of chasmogamous flowers: Seq-1-F (5'-AAGAACGAGGTTTGGCTTCA-3')/Seq-1-R (5'-AAC TTGACCCACATCAAGGA-3') and Seq-2-F (5'-GGAAAC ACAATCACTATGGCA-3')/Seq-2-R (5'-ATTGCATGTACA TGCCAGCTA-3'). The PCR was conducted using Platinum *Taq* DNA Polymerase High Fidelity (Invitrogen) following the manufacturer's instructions with some modifications (i.e.,



annealing temperature of 58°C and elongation time of 90 s). The PCR products were separated by electrophoresis on 1% agarose gels stained with ethidium bromide. The DNA fragments of the expected size were purified using the E.Z.N.A Gel Extraction Kit (Omega Bio-tek), and sequenced using the 3730xl DNA Analyzer (Applied Biosystems). The deduced Vr06g12650 protein and 11 *A. thaliana* YUC proteins (i.e., AT4G32540.1, AT4G13260.1, AT1G04610.1, AT5G11320.1, AT5G43890.1, AT5G25620.2, AT2G33230.1, AT4G28720.1, AT1G04180.1, and AT1G48910.1) obtained from The *Arabidopsis* Information Resource² database were used to construct a phylogenetic tree using the “One Click” mode and LRT statistical test of Phylogeny.fr (Dereeper et al., 2010).

RESULTS

Morphological Features of the Chasmogamous Mutant Mungbean and Inheritance of the Floral Trait

The floral architecture of CM plants differed from that of Sulu-1 plants. The CM flowers were missing the wing and keel petals (Figure 1). When the CM and Sulu-1 plants were crossed, the F₁ hybrids had normal flowers (Figure 1), which suggested a recessive gene (or genes) regulated the production of chasmogamous flowers in CM plants.

We determined the segregation ratio of the chasmogamous flower trait in F₂ plants. Out of 127 plants, 96 had normal flowers, while 31 had chasmogamous flowers. This segregation corresponded with a 3:1 ratio ($\chi^2 = 0.02$, $P = 0.88$), indicating that a single recessive gene mediated the chasmogamous flower trait in CM plants. This confirmed the results of the study by Sorajjapinun and Srinives (2011).

Development of Indel Markers

To map the gene responsible for the production of chasmogamous flowers, indel markers were developed using the

transcriptomes of the Sulu-1 and CM parents. We randomly selected 140 transcript sequences carrying indels to develop markers, and determined that 84 of them (i.e., 57.1%) were polymorphic between the parents (Supplementary Table S1). All of the markers were co-dominant, and able to distinguish between the parents and hybrid progenies. Seventy-four markers were localized to 11 chromosomes by aligning their related transcripts with the whole mungbean genome sequence (Figure 2). Another 10 markers were located on scaffolds that could not be assembled on chromosomes.

Genetic Mapping of the *cha* Gene

Linkage analysis of the polymorphic indel markers and phenotyping of the flowers were completed using 127 plants from the F₂ population. The target *cha* gene was mapped to chromosome 6 between markers VRID155 and VRID120 at a distance of 1.6 and 3.4 cM, respectively (Figure 3A). To locate *cha* more precisely, seven polymorphic SSR markers were developed in the target interval by screening SSR motifs in the reference genome sequence (Supplementary Table S2). Eight recombinants were identified in the interval between VRID115 and VRID120. By associating the marker genotypes with floral phenotypes, the 60 and 38 recombinants restricted *cha* a segment between markers SSR09 and SSR12. These two markers were 277.1 kb apart, and were located at 30.40 and 30.68 Mb of chromosome 6, respectively (Figure 3B). Based on the mungbean whole genome sequence, 12 candidate genes (i.e., *Vradi06g12620* to *Vradi06g12730*; Supplementary Table S3) were detected in this region.

Function Prediction and Sequencing of *cha* Candidate Genes

To predict the functions of the 12 detected *cha* candidate genes, the putative protein sequences encoded by these genes were used as queries to search the NCBI database. Their predicted functions are listed in Table S3. *Vradi06g12650*, which encoded a YUC homolog (Figure 3) related to the auxin biosynthesis pathway and floral development, was

²<http://arabidopsis.org>



considered the most likely *cha* gene. A BLASTP search revealed that the YUC protein encoded by *Vradi06g12650* was most similar to the *A. thaliana* proteins AtYUC4 (identity: 68.13%) and AtYUC1 (identity: 62.32%). A protein sequence-based phylogenetic analysis involving *Vradi06g12650* and 11 *A. thaliana* YUC proteins also revealed a close relationship between *Vradi06g12650* and AtYUC1/AtYUC4 (Figure 4). We sequenced the *Vradi06g12650* coding domain in Sulu-1, CM, and wild-type V1197 mungbean plants. Comparisons among the resulting sequences and the corresponding reference sequence (i.e., from VC1973A) revealed six single nucleotide

polymorphisms (SNPs) between the CM and Sulu-1 genes (Figure 5). The SNP at the 382-bp position led to an amino acid substitution from glutamine to glutamic acid, while the other SNPs resulted in non-sense mutations (Figures 5 and 6). The six SNPs were not detected between the CM and V1197 sequences. However, a 1-bp deletion at the 894-bp position was observed in the CM *Vradi06g12650* sequence (Figure 5). This deletion produced a frame-shift mutation in the 3'-terminus, resulting in a shorter CM *Vradi06g12650* protein with a different C-terminus compared with the corresponding wild-type V1197, Sulu-1, and VC1973A proteins (Figure 6). Additionally, this deletion was

detected in a sequence (i.e., Unigene0038420) from the CM transcriptome (Supplementary Data 1). These results suggested that the 1-bp deletion in *Vradi06g12650* is responsible for the production of chasmogamous flowers in CM plants.

DISCUSSION

Advantages of Developing Markers from Transcriptomes and the Utility of *cha* for Mungbean Improvement

Mungbean is not a model plant. Therefore, its molecular genetics and genome have not been as comprehensively studied as in many other crops. The available molecular markers are insufficient for genetics research and mungbean breeding programs. Additionally, some reports have indicated that the polymorphic information content of mungbean SSR markers is low (Tangphatsornruang et al., 2009; Chen et al., 2015). Thus, it is important to develop more mungbean genomic resources. In this study, we developed indel markers by comparing the transcriptomes of two mungbean accessions, which resulted in the detection of several polymorphic markers between the parents of the mapping population. The differences in indel sizes between Sulu-1 and CM plants enabled the indels to be distinguished by PCR followed by 3% agarose gel electrophoresis analysis (Supplementary Figure S1), which is easier to complete than polyacrylamide gel electrophoresis. This resulted in a more efficient genetic mapping procedure.

With the newly developed markers, the *cha* gene responsible for the production of chasmogamous flowers in mungbean was mapped to chromosome 6. Several newly developed markers were identified closely linked to the target gene. Our data revealed that the marker SSR10 was completely linked to *cha*. However, this finding is based on a relatively small population of 127 F₂ plants. Because the CM floral phenotype was observed to be regulated by a single recessive gene, these linked markers may be useful for the marker-assisted selection of mungbean plants producing chasmogamous flowers.

Functions of the *cha* Candidate Gene

Vradi06g12650 is the most likely *cha* gene. This gene putatively encodes a YUC homolog involved in auxin biosynthesis and floral development (Zhao et al., 2001; Cheng et al., 2006). The YUC proteins constitute a family of FMOs containing several conserved sequence motifs, including the FAD-binding motif, FMO-identifying motif, NADPH-binding motif, and F/LATGY motif (Schlaich, 2007). The protein encoded by *Vradi06g12650* is highly similar to *A. thaliana* YUC4 (Supplementary Figure S2). The CM *Vradi06g12650* coding sequence differed from that of the wild-type V1197 by a 1-bp deletion at the 894-bp position. This deletion was only detected in CM plants, and results in a frame-shift of the coding sequence leading to the absence of the LATGY motif in the C-terminus of the predicted YUC protein (Figure 6). There are 11 *A. thaliana* genes encoding YUC proteins, suggesting there may be some functional redundancy among these proteins. Mutational inactivation of a single YUC

family gene in *A. thaliana* caused no obvious developmental defects, while a double *yuc1 yuc4* mutant and a quadruple *yuc1 yuc4 yuc10 yuc11* mutant exhibited severe defects in the formation of floral organs (Zhao et al., 2001; Cheng et al., 2006). These findings differed from our observation that a defect in a single YUC4-like gene causes a dramatic morphological abnormality in mungbean floral organs. However, it is worth noting that the mutation detected in our study produced an immature YUC4-like protein lacking the F/LATGY motif, which is highly conserved among YUC proteins (Supplementary Figure S2). Therefore, this defective YUC4-like protein may be responsible for the abnormal floral development in CM plants. Additional studies are required to characterize how *cha* affects the production of chasmogamous flowers in mungbean.

AUTHOR CONTRIBUTIONS

JC designed the InDel and SSR markers and prepared the manuscript. PS performed gene mapping and reviewed the manuscript. XChen involved in bioinformatics analysis and reviewed the manuscript. XCui sequenced the candidate gene. XY conducted the hybrids and developed the populations. PS designed the study and refined the manuscript.

FUNDING

This study was financially supported by grants from the National Natural Science Foundation of P. R. China (Grant no. 31271786) and Jiangsu Planned Projects for Postdoctoral Research Funds (Grant no. 1302034B).

ACKNOWLEDGMENTS

We are thankful to the Center for Agricultural Biotechnology, Kasetsart University, Joint Legume Research Center Kasetsart University and Jiangsu Academy of Agricultural Sciences, and Institute of Vegetable Crops, Jiangsu Academy of Agricultural Sciences/Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, for molecular lab facilities.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00830>

FIGURE S1| Polymorphic InDel markers on 3% agarose gel. Polymorphisms of 48 markers with large variance can be detected by 3% agarose gel. For each marker, bands from left to right are the PCR product from Sulu-1, CM and an F1 hybrid.

FIGURE S2| Protein sequence alignment of arabidopsis YUC1, YUC4 and *Cha* candidate. Sequences of AtYUC1 (AT4G32540.1) and AtYUC4 (AT5G11320.1) were downloaded from NCBI. *Cha* is the deduced protein of mungbean Vr06g12650.1. Alignment was performed by ClustalX, consensus sequence was marked by asterisks. Conserved motifs of FMOs were mark on the top of sequences.

REFERENCES

- Alam, A. K. M. M., Somta, P., and Srinives, P. (2014). Identification and confirmation of quantitative trait loci controlling resistance to mungbean yellow mosaic disease in mungbean [*Vigna radiata* (L.) Wilczek]. *Mol. Breed.* 34, 1497–1506. doi: 10.1007/s11032-014-0133-0
- Alam, A. M., Somta, P., Mukhtadir, M., and Srinives, P. (2014). Quantitative trait loci associated with seed weight in mungbean (*Vigna radiata* (L.) Wilczek). *Kasetsart J. (Nat. Sci.)* 48, 197–204.
- Aloni, R., Aloni, E., Langhans, M., and Ullrich, C. I. (2006). Role of auxin in regulating *Arabidopsis* flower development. *Planta* 223, 315–328. doi: 10.1007/s00425-005-0088-9
- Bennett, S. R., Alvarez, J., Bossinger, G., and Smyth, D. R. (1995). Morphogenesis in pinoid mutants of *Arabidopsis thaliana*. *Plant J.* 8, 505–520. doi: 10.1046/j.1365-3113.1995.8040505.x
- Chankaew, A., Somta, P., Isemura, T., Tomooka, N., Kaga, A., Vaughan, D. A., et al. (2013). QTL mapping reveals conservation of major and minor loci for powdery mildew resistance in four sources of resistance in mungbean [*Vigna radiata* (L.) Wilczek]. *Mol. Breed.* 32, 121–130. doi: 10.1007/s11032-013-9856-6
- Chankaew, S., Somta, P., Sorajjapinun, W., and Srinives, P. (2011). Quantitative trait loci mapping of Cercospora leaf spot resistance in mungbean, *Vigna radiata* (L.) Wilczek. *Mol. Breed.* 28, 255–264.
- Chen, H., Qiao, L., Wang, L., Wang, S., Blair, M. W., and Cheng, X. (2015). Assessment of genetic diversity and population structure of mung bean (*Vigna radiata*) germplasm using EST-based and genomic SSR markers. *Gene* 566, 175–183. doi: 10.1016/j.gene.2015.04.043
- Chen, X., Sorajjapinun, W., Reiwithongchum, S., and Srinives, P. (2003). Identification of parental mungbean lines for production of hybrid varieties. *CMU J.* 2, 97–106.
- Cheng, Y., Dai, X., and Zhao, Y. (2006). Auxin biosynthesis by the YUCCA flavin monooxygenases controls the formation of floral organs and vascular tissues in *Arabidopsis*. *Genes Dev.* 20, 1790–1799. doi: 10.1101/gad.1415106
- Cheng, Y., Dai, X., and Zhao, Y. (2007). Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *Plant Cell* 19, 2430–2439. doi: 10.1105/tpc.107.053009
- Cheng, Y., and Zhao, Y. (2007). A role for auxin in flower development. *J. Integr. Plant Biol.* 49, 99–104. doi: 10.1111/j.1744-7909.2006.00412.x
- Cherian, C. A., Mallikarjuna, N., Jadhav, D., and Saxena, K. (2006). Open flower segregants selected from *Cajanus platycarpus* crosses. *J. SAT Agric. Res.* 2, 1–2.
- Chotechung, S., Somta, P., Chen, J., Yimram, T., Chen, X., and Srinives, P. (2016). A gene encoding a polygalacturonase-inhibiting protein (PGIP) is a candidate gene for bruchid (Coleoptera: Bruchidae) resistance in mungbean (*Vigna radiata*). *Theor. Appl. Genet.* doi: 10.1007/s00122-016-2731-1 [Epub ahead of print].
- Dahiya, B., Lather, V., Solanki, I., and Kumar, R. (1984). Genetic resources: useful spontaneous mutants in chickpea. *Int. Chick. News.* 11, 5–7.
- Dereeper, A., Audic, S., Claverie, J. M., and Blanc, G. (2010). BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol. Biol.* 10:8. doi: 10.1186/1471-2148-10-8
- Dheer, M., Punia, S. S., Ram, B., Jain, N. K., Koli, N. R., Bhatnagar, V., et al. (2014). Morphological features of an open flower mutant plant and characterization of their progenies in lentil (*Lens culinaris* Medik.). *Genet. Resour. Crop Evol.* 61, 879–886. doi: 10.1007/s10722-014-0100-y
- Galweiler, L., Guan, C., Muller, A., Wisman, E., Mendgen, K., Yephremov, A., et al. (1998). Regulation of polar auxin transport by AtPIN1 in *Arabidopsis* vascular tissue. *Science* 282, 2226–2230. doi: 10.1126/science.282.5397.2226
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–652. doi: 10.1038/nbt.1883
- Gupta, S. K., Bansal, R., and Gopalakrishna, T. (2014). Development and characterization of genic SSR markers for mungbean (*Vigna radiata* (L.) Wilczek). *Euphytica* 195, 245–258.
- Gwag, J. G., Chung, J. W., Chung, H. K., Lee, J. H., Ma, K. H., Dixit, A., et al. (2007). Characterization of new microsatellite markers in mung bean, *Vigna radiata* (L.). *Mol. Ecol. Resour.* 6, 1132–1134. doi: 10.1007/s11033-014-3436-7
- Isemura, I., Kaga, A., Tabata, S., Somta, P., Srinives, P., Shimizu, T., et al. (2012). Construction of a genetic linkage map and genetic analysis of domestication related traits in mungbean (*Vigna radiata*). *PLoS ONE* 7:e41304. doi: 10.1371/journal.pone.0041304
- Kajonphol, T., Sangsiri, C., Somta, P., Toojinda, T., and Srinives, P. (2012). SSR map construction and quantitative trait loci (QTL) identification of major agronomic traits in mungbean (*Vigna radiata* (L.) Wilczek). *SABRAO J. Breed. Genet.* 44, 71–86.
- Kang, Y. J., Kim, S. K., Kim, M. Y., Lestari, P., Kim, K. H., Ha, B. K., et al. (2014). Genome sequence of mungbean and insights into evolution within *Vigna* species. *Nat. Commun.* 5, 5543. doi: 10.1038/ncomms6443
- Kasettranan, W., Somta, P., and Srinives, P. (2010). Mapping of quantitative trait loci controlling powdery mildew resistance in mungbean (*Vigna radiata* (L.) Wilczek). *J. Crop Sci. Biotechnol.* 13, 155–161. doi: 10.1007/s12892-010-0052-z
- Kato, H., and Namai, H. (1987). Floral characteristics and environmental factors for increasing natural outcrossing rate for F1 hybrid seed production of rice *Oryza sativa* L. *Jpn. J. Breed.* 37, 318–330. doi: 10.1270/jsbbs1951.37.318
- Kitsanachandee, R., Somta, P., Chatchawankanphanich, O., Akhtar, K. P., Shah, T. M., Nair, R. M., et al. (2013). Detection of quantitative trait loci for mungbean yellow mosaic India virus (MYMIV) resistance in mungbean (*Vigna radiata* (L.) Wilczek) in India and Pakistan. *Breed. Sci.* 63, 367–373. doi: 10.1270/jsbbs.63.367
- Krizek, B. A., and Fletcher, J. C. (2005). Molecular mechanisms of flower development: an armchair guide. *Nat. Rev. Genet.* 6, 688–698. doi: 10.1038/nrg1675
- Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, M. J., Lincoln, S. E., et al. (1987). MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1, 174–181. doi: 10.1016/0888-7543(87)90010-3
- Li, Q., and Wan, J.-M. (2005). SSRHunter: development of a local searching software for SSR sites. *Hereditas* 27, 808–810.
- Lodhi, M. A., Ye, G.-N., Weeden, N. F., and Reisch, B. I. (1994). A simple and efficient method for DNA extraction from grapevine cultivars and *Vitis* species. *Plant Mol. Biol. Rep.* 12, 6–13. doi: 10.1007/BF02668658
- Mather, K. (1951). *The Measurement of Linkage in Heredity*. London: Methuen & Co. LTD.
- Okada, K., Ueda, J., Komaki, M. K., Bell, C. J., and Shimura, Y. (1991). Requirement of the auxin polar transport system in early stages of *Arabidopsis* floral bud formation. *Plant Cell* 3, 677–684. doi: 10.1105/tpc.3.7.677
- Prathet, P., Somta, P., and Srinives, P. (2012). Mapping QTL conferring resistance to iron deficiency chlorosis in mungbean [*Vigna radiata* (L.) Wilczek]. *Field Crops Res.* 137, 230–236. doi: 10.1016/j.fcr.2012.08.002
- Przemek, G. K., Mattsson, J., Hardtke, C. S., Sung, Z. R., and Berleth, T. (1996). Studies on the role of the *Arabidopsis* gene MONOPTEROS in vascular development and plant cell axialization. *Planta* 200, 229–237. doi: 10.1007/BF00208313
- Pundir, R., and Reddy, G. (1998). Two new traits-open flower and small leaf in chickpea (*Cicer arietinum* L.). *Euphytica* 102, 357–361. doi: 10.1023/A:1018307709788
- Sangiri, C., Kaga, A., Tomooka, N., Vaughan, D., and Srinives, P. (2007). Genetic diversity of the mungbean (*Vigna radiata*, Leguminosae) gene pool on the basis of microsatellite analysis. *Aust. J. Bot.* 55, 837–847. doi: 10.1071/Bt07105
- Schlaich, N. L. (2007). Flavonoid-containing monooxygenases in plants: looking beyond detox. *Trends Plant Sci.* 12, 412–418. doi: 10.1016/j.tplants.2007.08.009
- Soehendi, R., and Srinives, P. (2005). Significance of heterosis and heterobeltiosis in F1 hybrid of mungbean (*Vigna radiata* (L.) Wilczek) for hybrid seed production. *SABRAO J. Breed. Genet.* 37, 97–105.
- Sompong, U., Somta, P., Raboy, V., and Srinives, P. (2012). Mapping of quantitative trait loci for phytic acid and phosphorus contents in seed and seedling of mungbean (*Vigna radiata* (L.) Wilczek). *Breed. Sci.* 62, 87–92. doi: 10.1270/jsbbs.62.87
- Somta, P., Musch, W., Kongsamai, B., Chanprame, S., Nakasathien, S., Toojinda, T., et al. (2008). New microsatellite markers isolated from mungbean (*Vigna radiata* (L.) Wilczek). *Mol. Ecol. Resour.* 8, 1155–1157. doi: 10.1111/j.1755-0998.2008.02219.x
- Somta, P., Seehalak, W., and Srinives, P. (2009). Development, characterization and cross-species amplification of mungbean (*Vigna radiata*) genic microsatellite markers. *Conserv. Genet.* 10, 1939–1943. doi: 10.1007/s10592-009-9860-x
- Sorajjapinun, W., Chankaew, S., and Srinives, P. (2012). Comparison of hybrid vigor based on parental distance in SSR markers and agronomic traits in

- mungbean (*Vigna radiata* (L.) Wilczek). *Sonklanakarin J. Sci. Technol.* 34, 143–151.
- Sorajjapinun, W., and Srinives, P. (2011). Chasmogamous mutant, a novel character enabling commercial hybrid seed production in mungbean. *Euphytica* 181, 217–222. doi: 10.1007/s10681-011-0408-z
- Srinivasan, S., and Gaur, P. M. (2012). Genetics and characterization of an open flower mutant in chickpea. *J. Hered.* 103, 297–302. doi: 10.1093/jhered/esr125
- Tangphatsornruang, S., Somta, P., Uthaisaisanwong, P., Chanprasert, J., Sangsrakru, D., Seehalak, W., et al. (2009). Characterization of microsatellites and gene contents from genome shotgun sequences of mungbean (*Vigna radiata* (L.) Wilczek). *BMC Plant Biol.* 9:137. doi: 10.1186/1471-2229-9-137
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., et al. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 40:e115. doi: 10.1093/nar/gks596
- Verdcourt, B. (1970). Studies in the Leguminosae-Papilionoideae for the 'Flora of Tropical East Africa': IV. *Kew Bull.* 24, 507–569. doi: 10.2307/4103051
- Voorrips, R. E. (2002). MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* 93, 77–78. doi: 10.1093/jhered/93.1.77
- Yashpal, Singh, M. N., Pathak, N., and Saroj, S. K. (2015). Combining ability, heterosis and inbreeding depression in inter specific hybrids involving greengram [*Vigna radiata* (L.) Wilczek] and blackgram [*Vigna mungo* (L.) Hepper]. *Electron. J. Plant Breed.* 6, 87–92.
- Zhao, Y., Christensen, S. K., Fankhauser, C., Cashman, J. R., Cohen, J. D., Weigel, D., et al. (2001). A role for flavin monooxygenase-like enzymes in auxin biosynthesis. *Science* 291, 306–309. doi: 10.1126/science.291.5502.306

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Chen, Somta, Chen, Cui, Yuan and Srinives. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Multiple QTL-Seq Strategy Delineates Potential Genomic Loci Governing Flowering Time in Chickpea

Rishi Srivastava^{1†}, Hari D. Upadhyaya^{2†}, Rajendra Kumar³, Anurag Daware¹,
Udita Basu¹, Philanim W. Shimray⁴, Shailesh Tripathi⁴, Chellapilla Bharadwaj⁴,
Akhilesh K. Tyagi^{1†} and Swarup K. Parida^{1*}

¹ National Institute of Plant Genome Research, New Delhi, India, ² International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India, ³ U.P. Council of Agricultural Research, Lucknow, India, ⁴ Division of Genetics, Indian Agricultural Research Institute, New Delhi, India

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University
of Mexico, Mexico

Reviewed by:

Rupesh Kailasrao Deshmukh,
Laval University, Canada
Matthew Nicholas Nelson,
Royal Botanic Gardens, Kew,
United Kingdom

*Correspondence:

Swarup K. Parida
swarup@nipgr.ac.in;
swarupdbt@gmail.com

† Present address:

Akhilesh K. Tyagi,
Department of Plant Molecular
Biology, University of Delhi South
Campus, New Delhi, India

†These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 19 December 2016

Accepted: 07 June 2017

Published: 11 July 2017

Citation:

Srivastava R, Upadhyaya HD,
Kumar R, Daware A, Basu U,
Shimray PW, Tripathi S,
Bharadwaj C, Tyagi AK and
Parida SK (2017) A Multiple QTL-Seq
Strategy Delineates Potential
Genomic Loci Governing Flowering
Time in Chickpea.
Front. Plant Sci. 8:1105.
doi: 10.3389/fpls.2017.01105

Identification of functionally relevant potential genomic loci using an economical, simpler and user-friendly genomics-assisted breeding strategy is vital for rapid genetic dissection of complex flowering time quantitative trait in chickpea. A high-throughput multiple QTL-seq strategy was employed in two inter (*Cicer arietinum* desi accession ICC 4958 × *C. reticulatum* wild accession ICC 17160)- and intra (ICC 4958 × *C. arietinum* kabuli accession ICC 8261)-specific RIL mapping populations to identify the major QTL genomic regions governing flowering time in chickpea. The whole genome resequencing discovered 1635117 and 592486 SNPs exhibiting differentiation between early- and late-flowering mapping parents and bulks, constituted by pooling the homozygous individuals of extreme flowering time phenotypic trait from each of two aforesaid RIL populations. The multiple QTL-seq analysis using these mined SNPs in two RIL mapping populations narrowed-down two longer (907.1 kb and 1.99 Mb) major flowering time QTL genomic regions into the high-resolution shorter (757.7 kb and 1.39 Mb) QTL intervals on chickpea chromosome 4. This essentially identified regulatory as well as coding (non-synonymous/synonymous) novel SNP allelic variants from two *efl1* (early flowering 1) and *G1* (*GIGANTEA*) genes regulating flowering time in chickpea. Interestingly, strong natural allelic diversity reduction (88–91%) of two known flowering genes especially mapped at major QTL intervals as compared to that of background genomic regions (where no flowering time QTLs were mapped; 61.8%) in cultivated vis-à-vis wild *Cicer* gene pools was evident inferring the significant impact of evolutionary bottlenecks on these loci during chickpea domestication. Higher association potential of coding non-synonymous and regulatory SNP alleles mined from *efl1* (36–49%) and *G1* (33–42%) flowering genes for early and late flowering time differentiation among chickpea accessions was evident. The robustness and validity of two functional allelic variants-containing genes localized at major flowering time QTLs was apparent by their identification from multiple intra-/inter-specific mapping populations of chickpea. The functionally relevant molecular tags delineated can be of immense use for deciphering the natural allelic diversity-based domestication pattern of flowering time and expediting genomics-aided crop improvement to develop early flowering cultivars of chickpea.

Keywords: chickpea, flowering time, multiple QTL-seq, QTL, SNP

INTRODUCTION

Chickpea (*Cicer arietinum* L.) is one of the vital food legume crops represented by two of its major cultivar types- *desi* and *kabuli* (Kumar et al., 2011) which are thought to be domesticated along with the wild ancestor *C. reticulatum* at Fertile Crescent around 10000 years ago (Burger et al., 2008; Toker, 2009). Development of early-flowering/maturing stress tolerant cultivars with high seed and pod yield is the prime objective of the present genomics-assisted breeding research in chickpea (Hegde, 2010; Zaman-Allah et al., 2011). The number of days to flowering is a major seed and pod-yield component trait that highly acclimatizes with climate change, diverse environmental and a/biotic stress factors and photoperiod response along with various other growth/developmental-related traits (Aryamanesh et al., 2010; Kashiwagi et al., 2013; Daba et al., 2016). Therefore, implication of flowering time in defining productivity as well as developing stress tolerant cultivars is well-documented in chickpea. Collectively, this infers that flowering time is a complex quantitative trait and it is governed by multiple major as well as minor genes/QTLs (quantitative trait loci). A strong impact of a known major evolutionary bottleneck- vernalization- on flowering time response during chickpea domestication infers that the flowering time is a most important domestication trait selected during breeding of presently cultivated *desi* and *kabuli* accessions (Abbo et al., 2014). The genetic dissection of complex flowering time quantitative trait by identifying the functionally relevant potential genes/alleles colocalized at QTLs governing this major yield component and domestication trait is thus imperative for their broader effective practical applicability in marker-aided genetic improvement of chickpea.

Significant progress has been made to decipher the complex genetic inheritance characteristics and molecular genetic dissection of flowering time trait in chickpea (Anbessa et al., 2006; Cobos et al., 2007; Pierre et al., 2008, 2011; Aryamanesh et al., 2010; Zhang et al., 2013). This involves identification of four different major early flowering (*efl*) gene loci/allelic variants [*efl1*, *efl2/ppd* (photoperiod), *efl3* and *efl4*] controlling varied flowering time trait adaptation characteristics in multiple chickpea accessions (Hegde, 2010; Gaur et al., 2014; Weller and Martínez, 2015). Additionally, this includes colocalization of various known flowering time gene homologs [like *Efl1*, *Efl2*, *LFY* (LEAFY) and *FT* (flowering time) gene families] within the low-resolution major flowering time QTL regions mapped on chickpea chromosomes (Cho et al., 2002; Anbessa et al., 2006; Lichtenzweig et al., 2006; Cobos et al., 2007, 2009; Aryamanesh et al., 2010; Hossain et al., 2010; Rehman et al., 2011; Vadez et al., 2012; Jamalabadi et al., 2013; Varshney et al., 2014).

Substantial efforts have also been made to understand the complex gene regulatory networks and transcriptional modules governing flower development in a *desi* chickpea accession (ICC 4958) through NGS (next-generation sequencing)-based global transcriptome sequencing strategy (Singh et al., 2013). The deployment of these differentially expressed candidate gene-derived SNPs in association mapping and their subsequent integration with GWAS (genome-wide association study),

high-resolution QTL mapping, differential transcript profiling, molecular haplotyping have delineated tissue/stage (flower bud/flower)-specific differentially regulated potential candidate genes underlying major QTLs regulating flowering time at a whole genome level in chickpea (Das et al., 2015b; Upadhyaya et al., 2015). Until yet, none of these identified genes harboring major flowering time QTLs have been validated in multiple genetic backgrounds (mapping populations) and identified through map-based cloning that could be employed for marker-aided genetic enhancement of chickpea. This could be restrained due to low marker genetic polymorphism particularly between parents of multiple intra-/inter-specific mapping populations along with limited accessibility of large size mapping populations and high-density genetic linkage maps of chickpea. An alternative genome-wide approach is thus essential for quick identification and molecular mapping (fine-mapping/map-based isolation) of high-resolution flowering time QTLs/genes in order to drive genomics-led crop improvement in chickpea.

For genetic mapping of major flowering time QTLs, conventional QTL mapping approach that primarily involves genotyping of large-scale SSR (simple sequence repeat) and SNP (single nucleotide polymorphism) markers among mapping individuals of diverse inter-/intra-specific populations is found much expedient in chickpea (Anbessa et al., 2006; Lichtenzweig et al., 2006; Cobos et al., 2007, 2009; Aryamanesh et al., 2010; Hossain et al., 2010; Gowda et al., 2011; Rehman et al., 2011; Vadez et al., 2012; Jamalabadi et al., 2013; Stephens et al., 2014; Varshney et al., 2014). This approach essentially identified a diverse array of low-resolution longer marker confidence interval spanning major QTLs associated with chickpea flowering and maturation time (Cho et al., 2002; Anbessa et al., 2006; Lichtenzweig et al., 2006; Cobos et al., 2007, 2009; Aryamanesh et al., 2010; Hossain et al., 2010; Rehman et al., 2011; Vadez et al., 2012; Jamalabadi et al., 2013; Varshney et al., 2014).

The freely accessible draft genome sequences are found much proficient to accelerate genome and transcriptome resequencing of diverse *desi*, *kabuli* and wild accessions that are most commonly utilized as parents for generating diverse intra- and inter-specific mapping populations of chickpea (Jain et al., 2013; Varshney et al., 2013; Parween et al., 2015). Aside genomic resources, multiple genetic resources including advanced generation recombinant inbred lines (RILs) and back-cross mapping populations as well as core/mini-core germplasm accessions exhibiting a broader range of phenotypic variation for flowering time trait are now available in chickpea (Upadhyaya et al., 2001, 2008; Gaur et al., 2014). All these available genetic and genomic resources essentially have assisted in utilization of a high-throughput NGS-based QTL-seq strategy vis-à-vis a commonly adopted traditional QTL mapping approach for fast genome-wide scanning and genetic mapping of major QTLs controlling various quantitative agronomic traits (for instance, 100-seed weight, pod number and root/total plant dry weight ratio) in chickpea (Das et al., 2015a, 2016; Singh et al., 2016).

To complement this, a multiple QTL-seq assay that relies on QTL-seq analysis in multiple mapping populations generated by inter-crossing of common parental accessions, has been

developed currently as a most promising genome-wide strategy for QTL mapping at a high-resolution scale (Das et al., 2016). Essentially, multiple QTL-seq involves whole genome NGS resequencing of DNA bulks (exhibiting two utmost contrasting phenotypic traits) constituted from homozygous individuals of multiple mapping populations comprising at least single common parent. This approach is found most promising based on its potential to validate QTL-seq-derived major QTLs identified from individual preliminary as well as advanced generation intra-/inter-specific mapping population in multiple mapping populations of diverse genetic backgrounds. Moreover, utility of this approach is clearly evident from its efficacy to narrow-down each QTL-seq originated sizeable long QTL genomic intervals into functionally relevant potential candidate genes governing important agronomic traits (for instance, pod number) in chickpea (Das et al., 2016).

Considering usefulness and broader practical applicability, multiple QTL-seq assay can be employed for rapid genome-wide scanning and fine-mapping (positional cloning) of trait-linked major genes and natural allelic-variants colocalized at robust QTLs (well-validated in multiple mapping populations) in chickpea with minimal resource expenses. This will collectively enrich our understanding on complex genetic architecture and evolutionary pattern influencing flowering time quantitative trait variation during domestication of chickpea in order to expedite its genomics-assisted crop improvement. In view of afore-mentioned possibilities, a multiple QTL-seq strategy was employed in two inter- and intra-specific RIL (recombinant inbred lines) mapping populations- (*C. arietinum desi* accession ICC 4958 \times *C. reticulatum* wild accession ICC 17163) and (ICC 4958 \times *C. arietinum kabuli* accession ICC 8261)- at a genome-wide scale to delineate major genomic (gene) regions and novel natural allelic variants underlying the QTLs associated with flowering time in chickpea.

MATERIALS AND METHODS

Development and Phenotyping of RIL Mapping Populations for Flowering Time

Two inter- and intra-specific F_9 RIL mapping populations- (ICC 4958 \times ICC 17163, population size: 260) and (ICC 4958 \times ICC 8261, 204)- with contrasting flowering time trait were developed by single seed descent method. As per field phenotyping at International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), ICC 4958 (traditional cultivar/landrace, originated from India) is an early flowering chickpea accession with DTF (days to 50% flowering time) of 43 days. In contrast, ICC 17163 (wild accession) and ICC 8261 (traditional cultivar/landrace) originated from Turkey are late flowering chickpea accessions with DTF of 85 and 61 days, respectively. The *desi* chickpea accession ICC 4958 was considered as a common parent for both mapping populations generated.

For phenotyping, the mapping individuals and parents of both RIL populations were grown and phenotyped in the field as per RCBD (randomized complete block design) with two replications at two diverse eco-geographical regions [(ICRISAT,

Patancheru, Hyderabad: latitude 17° 3' N/longitude 77° 2' E from October to February) and (National Institute of Plant Genome Research (NIPGR), New Delhi: 28° 4' N/77° 2' E from November to March)] of India for two successive years (2013 and 2014) during crop growing season. In addition, these parents and RIL individuals were grown in the greenhouse to determine the flowering time response of these mapping individuals under both long- and short-day conditions at $22 \pm 2^\circ\text{C}$ following Upadhyaya et al. (2015). Ten to fifteen representative plants were screened from each mapping individual and parental accession of two RIL populations, and DTF of each individual/accession was calculated as per Upadhyaya et al. (2015). The homogeneity of RIL mapping populations across two locations/years as well as major parameters contributing to genetic inheritance characteristics such as frequency distribution, CV (coefficient of variation), H^2 (broad-sense heritability) of DTF trait among RIL individuals were determined as per Bajaj et al. (2015b). To evaluate the genetic inheritance pattern of flowering time trait, the interactions of mapping individuals/parents (G) with their phenotyping environments (E; like years and locations) were calculated using ANOVA (analysis of variance).

Whole Genome Resequencing and Multiple QTL-Seq Analysis

We selected 10 of each early and late flowering homozygous mapping individuals belonging to two extreme ends of DTF normal frequency distribution curve from each of the two RIL populations of ICC 4958 \times ICC 17163 and ICC 4958 \times ICC 8261 for QTL-seq study. Prior to inclusion of these selected 20 RIL mapping individuals in QTL-seq analysis, the homozygous genetic constitution of these individuals from both RIL populations for either of the early and late flowering trait was assured using their DTF field phenotyping data and genome-wide SSR markers-based genotyping information as per Das et al. (2015a, 2016).

The genomic DNA was isolated from constituted DNA bulks- early days to 50% flowering time bulk (EDTFB) and late days to 50% flowering time bulk (LDTFB) as well as parents of mapping populations using QIAGEN DNeasy kit (QIAGEN, United States) following manufacturer's instructions. The quantity and quality of isolated genomic DNA was ensured by Qubit 2.0 Fluorometer (Invitrogen Life Technologies, United States) and Bioanalyzer 2100 (Agilent Technologies, United States), respectively. About 1 μg of high-quality genomic DNA of each sample was utilized for library preparation using Illumina TruSeq DNA PCR-Free Library Preparation Kit according to the manufacturer's protocol. The libraries were processed for paired-end sequencing (100-nucleotide long reads) using Illumina HiSeq2000 platform (Illumina Technologies, United States) and raw sequence data were filtered through standard Illumina pipeline. The FASTQ sequences were further processed through NGS QC Toolkit v2.3 (Patel and Jain, 2012) to remove low-quality including primer/adaptor contaminated sequence reads. The filtered reads with a minimum phred Q-score of 30 across > 95% of nucleotide sequence were considered as high-quality.

Recently, a whole genome high-quality sequence assembly, including large size (510.9 Mb) chromosome pseudomolecule (334 Mb) and scaffolds of *desi* (ICC 4958) chickpea genome are freely available in public domain (Parween et al., 2015). Accordingly, a *desi* chickpea accession ICC 4958 for which the genome sequence is available was utilized as one of the common parent in two RIL mapping populations developed for our QTL-seq analysis. Therefore, we preferably utilized the latest released *desi* reference genome sequence as an anchor to mine resequencing-based SNPs from mapping parents and bulks for QTL-seq study at a whole genome level in chickpea.

High-quality sequence reads generated from parental accessions and bulks (EDTFB and LDTFB) were mapped onto reference *desi* chickpea genome using BWA with default parameters (Parween et al., 2015). Consequently, the uniquely mapped sequence reads were normalized in accordance with read coverage-depth among mapping parents and RIL individuals forming EDTFB and LDTFB bulks (Supplementary Table S1). The mined homozygous high-quality SNPs (minimum sequencing read-depth 10 with mean base quality ≥ 20) exhibiting differentiation between parents as well as between EDTFB and LDTFB were structurally and functionally annotated with respect to reference *desi* chickpea genome following Kujur et al. (2015a,b,c). As per the earlier defined recommended parameters of Takagi et al. (2013), Lu et al. (2014), and Das et al. (2015a, 2016), SNP-index and Δ (SNP-index)-led QTL-seq assay was employed in two RIL mapping populations individually to identify major DF QTLs in chickpea. The subtraction of SNP-index (percentage of SNPs-supporting sequence reads completely different from reference *desi* genome) between EDTFB and LDTFB was measured as Δ (SNP-index). Following the representation of genomic fragments obtained from ICC 4958 (early DTF) and ICC 17163/ICC 8261 (late DTF) in entire genome sequences, the SNP-index was calculated as “0” and “1,” respectively. The major genomic regions underlying QTL-seq derived DTF QTLs were ascertained by Δ (SNP-index) which is altogether different from 0 at a 99% significance level and thereby, considered to be highly significant QTLs governing flowering time in chickpea. A 10 Mb window-size and 1 kb increment sliding window approach was used to evaluate the mean distribution of Δ (SNP-index) of SNPs physically mapped across chromosomes in a target genomic interval. The SNP-index plots of EDTFB and LDTFB for two individual RIL mapping populations and null hypothesis statistical confidence intervals of Δ (SNP-index) were obtained to determine the accuracy and validity of QTL-seq derived QTLs following Takagi et al. (2013), Lu et al. (2014), and Das et al. (2015a, 2016).

Natural Allelic Diversity in Flowering Genes

The novel natural SNP allelic variants of flowering time-associated candidate genes underlying major DTF QTLs validated by multiple QTL-seq assay, were genotyped using the genomic DNA of 172 including 93 cultivated (39 *desi* and 53 *kabuli* accessions) and 79 wild chickpea accessions

[*C. reticulatum* (14 accessions), *C. echinospermum* (8), *C. judaicum* (22), *C. bijugum* (19) and *C. pinnatifidum* (15) and *C. microphyllum* (1); Bajaj et al. (2015a)] through Sequenom MALDI-TOF MassARRAY¹ as per Saxena et al. (2014a,b). The SNP allelic genotyping data generated among chickpea accessions were analyzed with TASSEL v5.0 (100-kb non-overlapping sliding window) to estimate the various nucleotide diversity parameters ($\theta\pi$ and Tajima's D) following Bajaj et al. (2015a) and Kujur et al. (2015a). For association analysis, the genotyping data of SNPs derived from flowering time-associated genes was integrated with DTF field and greenhouse-based phenotyping information, population structure ancestry coefficient (Q-matrix), kinship-matrix (K) and principal component analysis (PCA; P) data of 172 accessions following the detailed methods as described by Upadhyaya et al. (2015) to determine the SNP allele effect on early and late flowering time differentiation in chickpea.

RESULTS

Genetic Inheritance Pattern of Flowering Time in Mapping Populations

A significant difference of DTF trait was observed on phenotyping in field at two diverse geographical locations and in green house (long- and short-day photoperiod conditions) for 2 years. This varied from 25.8 to 99.5 days with 47.1 to 57.8 days mean \pm 13.5–16.1 days standard deviation (SD) of DTF trait with 25.6–32.1% CV and 79–86% H² among 260 individuals and parents of an inter-specific RIL mapping population [ICC 4958 (33.6–46.8 days mean \pm 2.1–3.7 days SD) \times ICC 17163 (85.1–91.5 days mean \pm 3.3–4.1 days SD)] (Table 1). A wider phenotypic variation (varied from 25.7 to 70.9 days with 46.1 to 53.9 days mean \pm 8.1–9.2 days SD) for DTF with 17.2–19.5% CV and 79–83% H² was detected among 204 individuals and parents of an another intra-specific RIL mapping population [ICC 4958 (32.5–49.7 mean \pm 2.7–3.6 SD) \times ICC 8261 (57.9–67.5 mean \pm 2.8–4.1 SD)] phenotyped similarly in field at two different geographical locations and in green house (long- and short-day photoperiod) for 2 years (Table 1). A significant ($P < 0.0001$) difference in DTF of individuals representing both RIL mapping populations grown under long- and short-day photoperiod conditions at green-house across 2 years was apparent. We observed a continuous variation-based normal frequency distribution along with a bi-directional transgressive segregation of DTF trait in these two RIL mapping populations (Figures 1A,B).

NGS-Based Whole Genome Resequencing to Generate Sequences for QTL-Seq Study

For QTL-seq study, we performed high-throughput whole genome NGS resequencing of early and late flowering parents as well as bulks- EDTFB (mean DTF: 22.7–28.6 days) and

¹<http://www.sequenom.com>

TABLE 1 | Diverse statistical measures-based DTF (days to 50% flowering time) trait variation determined in two intra- and inter-specific chickpea RIL mapping populations grown in field at two diverse geographical locations of India and in greenhouse (long- and short-day conditions) for 2 years.

Mapping populations	Geographical locations	Parental accessions			F ₉ RIL mapping individuals			
		Years	ICC 4958	ICC 17163	Mean ± SD; (days)	Range (days)	Coefficient of variation (CV%)	Broad-sense Heritability (H ² %)
			(Mean ± SD; days)	(Mean ± SD; days)				
ICC 4958 × ICC 17163 (260 F ₉ RILs)	Patancheru (Hyderabad)-field	2012	43.0 ± 2.5	85.4 ± 3.4	52.9 ± 15.3	27.1–97.2	28.9	85
		2013	43.2 ± 2.1	85.1 ± 3.5	52.7 ± 13.5	26.8–98.2	25.6	83
	New Delhi-field	2012	42.8 ± 2.6	85.5 ± 3.6	52.1 ± 14.7	26.1–96.3	28.2	86
		2013	43.1 ± 2.8	85.4 ± 3.9	53.4 ± 13.8	25.8–97.1	25.8	85
	Patancheru-green house (Long-day condition)	2012	47.5 ± 3.4	89.7 ± 4.1	56.1 ± 16.1	31.1–99.2	28.7	81
		2013	46.8 ± 3.7	90.2 ± 3.9	57.8 ± 15.5	30.6–99.5	26.8	80
	Patancheru-green house (Short-day condition)	2012	35.4 ± 3.7	90.7 ± 3.7	48.5 ± 15.4	22.4–89.7	31.8	79
		2013	33.6 ± 3.4	91.5 ± 3.3	47.1 ± 15.1	23.1–90.1	32.1	80
Mapping populations	Geographical locations	Parental accessions			F ₉ RIL mapping individuals			
		Years	ICC 4958	ICC 8261	Mean ± SD; (days)	Range (days)	Coefficient of variation (CV%)	Broad-sense Heritability (H ² %)
			(Mean ± SD; days)	(Mean ± SD; days)				
ICC 4958 × ICC 8261 (204 F ₉ RILs)	Patancheru (Hyderabad)-field	2012	43.1 ± 2.7	61.0 ± 3.1	46.1 ± 8.7	25.7–64.9	18.9	83
		2013	43.5 ± 2.9	61.5 ± 3.4	46.5 ± 8.9	26.3–65.9	19.1	81
	New Delhi-field	2012	43.1 ± 3.1	60.8 ± 3.2	46.7 ± 9.1	26.5–65.1	19.5	80
		2013	42.8 ± 3.3	61.5 ± 2.8	46.9 ± 8.8	26.8–65.4	18.8	82
	Patancheru-green house (Long-day condition)	2012	48.2 ± 2.9	66.4 ± 4.1	49.5 ± 8.1	29.1–70.9	16.4	80
		2013	49.7 ± 3.4	67.5 ± 3.9	48.9 ± 8.4	28.9–69.2	17.2	80
	Patancheru-green house (Short-day condition)	2012	32.5 ± 3.6	58.7 ± 3.5	51.8 ± 9.2	30.5–69.5	17.8	80
		2013	34.2 ± 3.1	57.9 ± 3.8	53.9 ± 9.0	31.8–70.2	16.7	79

LDTFB (63.0–95.1 days)- of two RIL mapping populations [(ICC 4958 × ICC 17163) and (ICC 4958 × ICC 8261)]. Considering the significant effect of long- and short-day photoperiod on DTF response in two RIL mapping populations across 2 years at green house, early and late flowering bulks- EDTFB (mean DTF: 25.3–30.1) and LDTFB (69.9–94.6) constituted from two RIL mapping populations were made at these two different photoperiod conditions separately and sequenced at a genome-wide scale for QTL-seq study. This produced 173.8 million high-quality average sequence reads (ranged from 167.4 to 181.4 million reads) with a ~30-fold sequencing-depth coverage. The sequencing data generated in the present study were submitted to NCBI-sequence read archive (SRA) database² with accession number SRR2229140 for unrestricted public access. About 86.5 to 90.1% sequence reads of these were mapped to unique physical locations of reference *desi* genome with a 73.6% mean coverage (Supplementary Table S1). To reduce the potential bias of read-depth in samples, the uniquely mapped sequence reads obtained from parents and bulks (EDTFB and LDTFB) of two RIL mapping populations were normalized in accordance with read coverage-depth. We measured the overall mapping efficiency of

non-redundant uniquely mapped sequence reads individually in mapping parents and bulks based on their sequencing-depth coverage (fold) as well as genome coverage (%) (Supplementary Table S1). This covered ~22.4-fold mean sequencing depth including 73.6% (544.7 Mb) of *desi* chickpea genome (estimated genome size ~740 Mb). For QTL-seq analysis, we compared the individual normalized sequence reads generated from mapping parents and bulks (EDTFB and LDTFB) with that of reference *desi* genomic sequences including pseudomolecules to discover homozygous high-quality SNPs.

Molecular Mapping of QTL-Seq Driven Major DF QTLs in a Mapping Population of ICC 4958 × ICC 17163

We discovered 1635117 SNPs (with an average map-density of 0.20 kb) revealing polymorphism between early (ICC 4958 and EDTFB) and late (ICC 17163 and LDTFB) flowering mapping parents and bulks according to their congruent physical positions (bp) on the reference pseudomolecule of *desi* genome (Table 2 and Supplementary Tables S2, S3). We measured the SNP-index of all individual SNPs exhibiting differentiation between early (ICC 4958 and EDTFB) and late (ICC 17163 and LDTFB)

²<http://www.ncbi.nlm.nih.gov/sra>

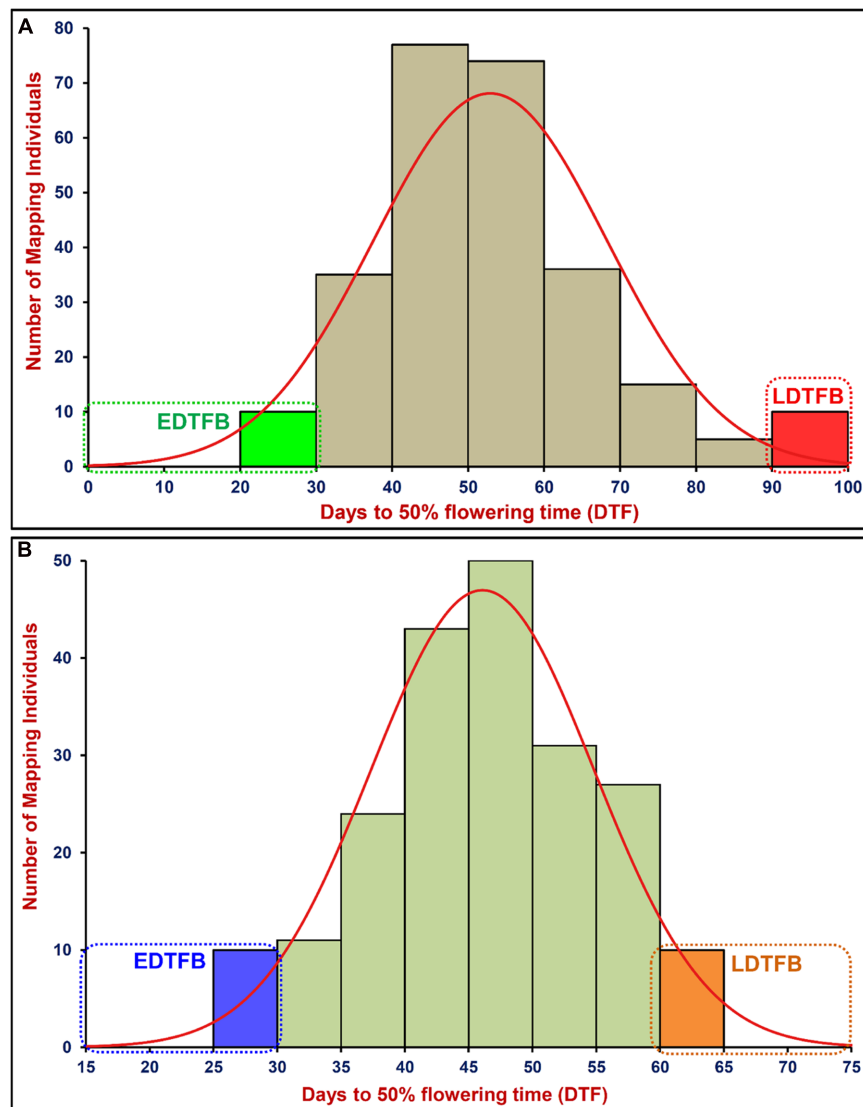


FIGURE 1 | Frequency distribution of mean days to 50% flowering time (DTF) trait variation measured among 260 and 204 mapping individuals and parents of two F_9 inter-/intra-specific RIL populations of ICC 4958 × ICC 17163 (A) and ICC 4958 × ICC 8261 (B), respectively. These were grown and phenotyped in field at two different geographical locations and green house (long- and short-day photoperiod) for 2 years. EDTFB/LDTFB: early/late days to 50% flowering time bulk.

flowering mapping parents and bulks, and plotted these SNP-index against chromosomes of reference genome. A 1-kb sliding window approach was employed to measure the mean SNP-index individually within a 10-Mb target genomic interval. Further, the Δ (SNP-index) was calculated by integrating the SNP-index of EDTFB and LDTFB, which were plotted across the genomic locations (Mb) of reference genome (Figure 2A).

We identified two major genomic regions (*Caq^aDTF4.1*: 46023168 to 46780835 bp and *Caq^aDTF4.2*: 26100745 to 28089632 bp) on chromosome 4 demonstrating the mean SNP-index of ≥ 0.8 in EDTFB and ≤ 0.2 in LDTFB in accordance with the SNP-index measurement criteria defined in QTL-seq analysis (Takagi et al., 2013; Lu et al., 2014; Das et al., 2015a, 2016; Figures 2A, 3A). The comprehensive analysis of these

selected DTF QTL genomic regions indicated the presence of majority of the SNP alleles derived from early (ICC 4958) and late (ICC 17163) flowering mapping parents in the early and late flowering mapping individuals composing the EDTFB and LDTFB bulks, respectively. Summarily, the QTL-seq assay in an inter-specific mapping population (ICC 4958 × ICC 17163) assured the occurrences of two major DTF QTLs- *Caq^aDTF4.1* and *Caq^aDTF4.2*- at the 1.99 Mb [26100745 (SNP_1A) to 28089632 (SNP_2A) bp with a Δ (SNP-index): 0.8] and 757.7 kb [46023168 (SNP_3A) to 46780835 (SNP_4A) bp with a Δ (SNP-index): 0.9] genomic intervals, respectively, on chickpea chromosome 4 (Figure 3A).

The detailed structural annotation of 16397 and 2542 SNPs at *Caq^aDTF4.1* and *Caq^aDTF4.2*, respectively, revealed

TABLE 2 | Genomic distribution of SNPs physically mapped on eight chromosomes of *desi* chickpea genome.

Chromosomes	Size (Mb) of <i>desi</i> chromosomes (pseudomolecules)	Number (%) of SNPs mapped		Average map density (kb)	
		<i>Desi</i> (ICC 4958) vs. Wild (ICC 17163)	<i>Desi</i> (ICC 4958) vs. <i>Kabuli</i> (CDC Frontier)	<i>Desi</i> (ICC 4958) vs. Wild (ICC 17163)	<i>Desi</i> (ICC 4958) vs. <i>Kabuli</i> (CDC Frontier)
<i>Ca_Desi_Chr01</i>	39.9	213381 (13.1)	96877 (16.4)	0.19	0.41
<i>Ca_Desi_Chr02</i>	33.2	170028 (10.4)	46869 (7.9)	0.20	0.71
<i>Ca_Desi_Chr03</i>	42.3	211478 (12.9)	43852 (7.4)	0.20	0.96
<i>Ca_Desi_Chr04</i>	55.0	298242 (18.2)	208734 (35.2)	0.18	0.26
<i>Ca_Desi_Chr05</i>	45.8	224449 (13.7)	40685 (6.9)	0.20	1.13
<i>Ca_Desi_Chr06</i>	54.8	264837 (16.2)	67588 (11.4)	0.21	0.81
<i>Ca_Desi_Chr07</i>	45.3	174065 (10.7)	72792 (12.3)	0.26	0.62
<i>Ca_Desi_Chr08</i>	17.7	78637 (4.8)	15089 (2.5)	0.23	1.17
Total	334	1635117	592486	0.20	0.56

the occurrence of 36.4 to 50% of SNPs in the genes and remaining in the intergenic regions (Supplementary Table S4). The gene-derived SNPs comprised of highest and lowest proportion of 72.7–73.5% and 1.1–1.3% SNPs in the DRRs (downstream regulatory regions) and URRs (upstream regulatory regions), respectively. The coding SNPs included the 45.8–54.5% synonymous and 45.5–54.2% non-synonymous (missense and nonsense) SNPs (Supplementary Table S4). The allelic variants of SNPs covering these major DTF QTLs (*Caq^aDTF4.1* and *Caq^aDTF4.2*) were further validated by resequencing of PCR fragments amplified from the parents (ICC 4958 and ICC 17163) and mapping individuals forming the EDTFB and LDFTB bulks. Accordingly, two major DTF QTLs were detected on similar aforesaid physical positions of chromosome 4 by the QTL-seq analysis of early- and late-flowering bulks (EDTFB and LDFTB) constituted from a RIL mapping population (ICC 4958 × ICC 17163) using the long- and short-day photoperiod-based greenhouse DTF phenotyping data of chickpea.

Molecular Mapping of QTL-Seq Driven Major DF QTLs in a Mapping Population of ICC 4958 × ICC 8261

A total of 592486 SNPs (with a mean map-density of 0.56 kb) were found polymorphic between early (ICC 4958 and EDTFB) and late (ICC 8261 and LDFTB) flowering mapping parents and bulks as per their congruent physical locations (bp) on the reference *desi* genome (pseudomolecule) (Table 2 and Supplementary Tables S2, S3). These genome resequencing-led SNPs were subsequently utilized for QTL-seq analysis.

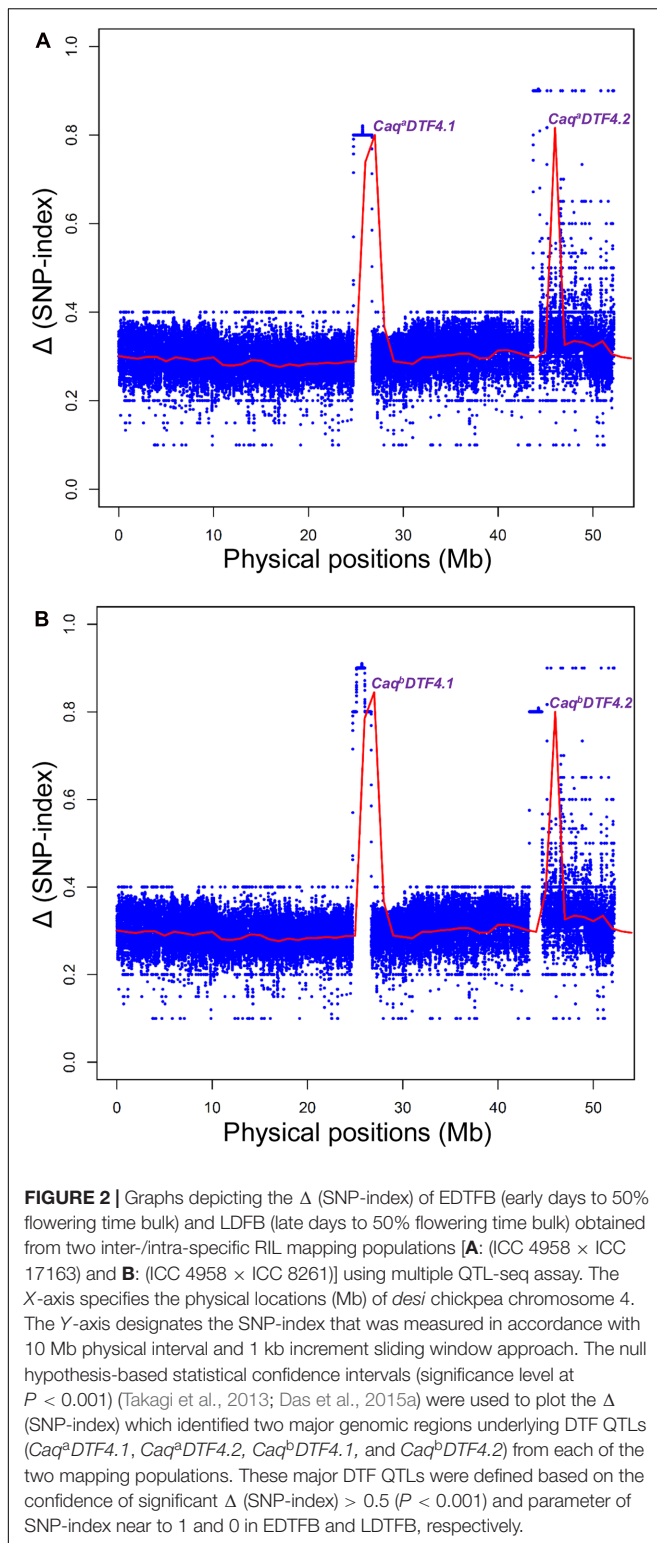
The SNP-index of individual SNPs exhibiting differentiation between early (ICC 4958 and EDTFB) and late (ICC 8261 and LDFTB) flowering mapping parents and bulks was estimated. The mean SNP-index (within a 1-kb sliding window and 10 Mb genomic interval) as well as Δ (SNP-index) of EDTFB and LDFTB were measured and further plotted across chromosomes as per aforesaid methods (Figure 2B). This essentially detected two major genomic regions (*Caq^bDTF4.1*: 45600294 to 46991993 bp and *Caq^bDTF4.2*: 26500027 to 27407090 bp) on chromosome 4 revealing the mean SNP-index

of ≥ 0.8 in EDTFB and ≤ 0.2 in LDFTB (Figures 2B, 3C,D). The accuracy of these major genomic regions underlying DTF QTLs was ascertained by a valid 99% Δ (SNP-index) significance level. The comprehensive analysis of these DTF QTL genomic regions inferred the occurrence of majority of the SNP alleles derived from parents (ICC 4958 and ICC 8261) in early and late flowering mapping individuals forming the EDTFB and LDFTB bulks, respectively. Overall, the QTL-seq in an intra-specific mapping population (ICC 4958 × ICC 8261) identified two major DTF QTLs- *Caq^bDTF4.1* and *Caq^bDTF4.2*- at the 907.1 kb [26500027 (SNP_1B) to 27407090 (SNP_2B) bp with a Δ (SNP-index): 0.9] and 1.39 Mb [45600294 (SNP_3B) to 46991993 (SNP_4B) bp with a Δ (SNP-index): 0.8] genomic intervals, respectively, on chickpea chromosome 4 (Figure 3B).

The detailed structural annotation of 7302 and 3177 SNPs at *Caq^bDTF4.1* and *Caq^bDTF4.2*, respectively, revealed the occurrence of 32.5 to 49.1% of SNPs in the genes and rest in the intergenic regions (Supplementary Table S4). The gene-derived SNPs included highest and lowest proportion of 73.8–74.6% and 1.2–1.5% SNPs in the DRRs and URRs, respectively. The coding SNPs included the 52.1–54.7% synonymous and 45.3–47.9% non-synonymous (missense and nonsense) SNPs (Supplementary Table S4). The allelic variants of SNPs covering the QTL-seq led major DTF QTLs (*Caq^bDTF4.1* and *Caq^bDTF4.2*) were validated by resequencing of PCR fragments amplified from the parents (ICC 4958 and ICC 8261) and mapping individuals composing the EDTFB and LDFTB bulks. Like-wise, we detected two major DTF QTLs on similar aforementioned physical positions of chromosome 4 by the QTL-seq analysis of EDTFB and LDFTB bulks constituted from a RIL mapping population (ICC 4958 × ICC 8261) using the long- and short-day photoperiod-based greenhouse DTF phenotyping data of chickpea.

Multiple QTL-Seq Rapidly Delineates Candidate Genes and Natural Allelic Variants Regulating Flowering Time in Chickpea

We correlated and compared the four major genomic regions underlying DTF QTLs (*Caq^aDTF4.1*, *Caq^aDTF4.2*, *Caq^bDTF4.1*



and $Caq^bDTF4.2$) identified and mapped by QTL-seq in two intra-/inter-specific RIL mapping populations of ICC 4958 \times ICC 17163 and ICC 4958 \times ICC 8261. Based on these analyses, two consensus major short physical genomic intervals of

757.7 kb [46023168 (SNP_3A) to 46780835 (SNP_4A) bp] and 907.1 kb [26500027 (SNP_1B) to 27407090 (SNP_2B) bp] harboring the major DTF QTLs were detected on chromosome 4 (Figures 3A,B). Our comprehensive multiple QTL-seq analysis in two inter-/intra-specific RIL mapping populations ascertained the validity of novel natural allelic variants-containing similar *efl1* and *GI* genes with a highest Δ (SNP-index) of 1.0 at $Caq^aDTF4.2$ and $Caq^bDTF4.1$ QTL regions governing flowering time in chickpea. Henceforth, these strong flowering time-associated *efl1* and *GI* genes localized at a major DTF QTL interval ($Caq^aDTF4.2$ and $Caq^bDTF4.1$) were considered as the potential candidates for flowering time regulation in chickpea. This essentially identified two upstream regulatory [46630632 (G/A) and 46630495 (C/T) bp] and one non-synonymous [Asparagine (AAT) to Serine (AGT)] coding [46618224 (A/G) bp] SNP allelic variants from a *efl1 desi* gene (Ca11444) as well as two synonymous coding [27092669 (G/A) bp] and downstream regulatory [27096726 (C/T) bp] SNP alleles from a *GI desi* gene (Ca10198) regulating flowering time in chickpea (Figures 3A,B and Supplementary Figure S1).

Natural Allelic Diversity-Led Domestication Pattern in Flowering Time Genes

The novel SNP allelic variants discovered from the coding (synonymous and non-synonymous) and non-coding regulatory sequence regions of two flowering genes, *efl1* (117 SNPs) and *GI* (31) localized at two major DTF QTL regions (identified by multiple QTL-seq) were genotyped in 93 *desi* and *kabuli* cultivated and 79 wild chickpea accessions to determine their natural/functional allelic diversity-based domestication pattern based on multiple nucleotide diversity parameters ($\theta\pi$ and Tajima's D) (Supplementary Table S5). The coding and regulatory SNPs discovered from the *efl1* (36–49% phenotypic variation explained) and *GI* (33–42%) flowering genes exhibited significantly higher association potential for early and late DTF differentiation among chickpea accessions (Table 3). Notably, only 9 to 12% of natural allelic variation-based functional diversity level estimated in *efl1* and *GI* flowering genes among wild gene pool was retained and thus got preserved in cultivated chickpea. The relative mean natural allelic diversity of two flowering genes (*efl1* and *GI*) localized at major DTF QTL regions between cultivated and wild chickpea varied from 88 to 91% ($\theta\pi_{Cc}/\theta\pi_{Cw}$). This was much higher than the relative mean natural allelic diversity level ($\theta\pi_{Cc}/\theta\pi_{Cw}$: 61.8%) estimated by using 7116 genome-wide SNPs localized at genomic regions where no DTF QTLs were mapped.

DISCUSSION

A broader phenotypic variation coupled with bi-directional transgressive segregation (normal frequency distribution) of DTF trait among RIL individuals and parents of inter (ICC 4958 \times ICC 17163)- and intra (ICC 4958 \times ICC 8261)-specific mapping

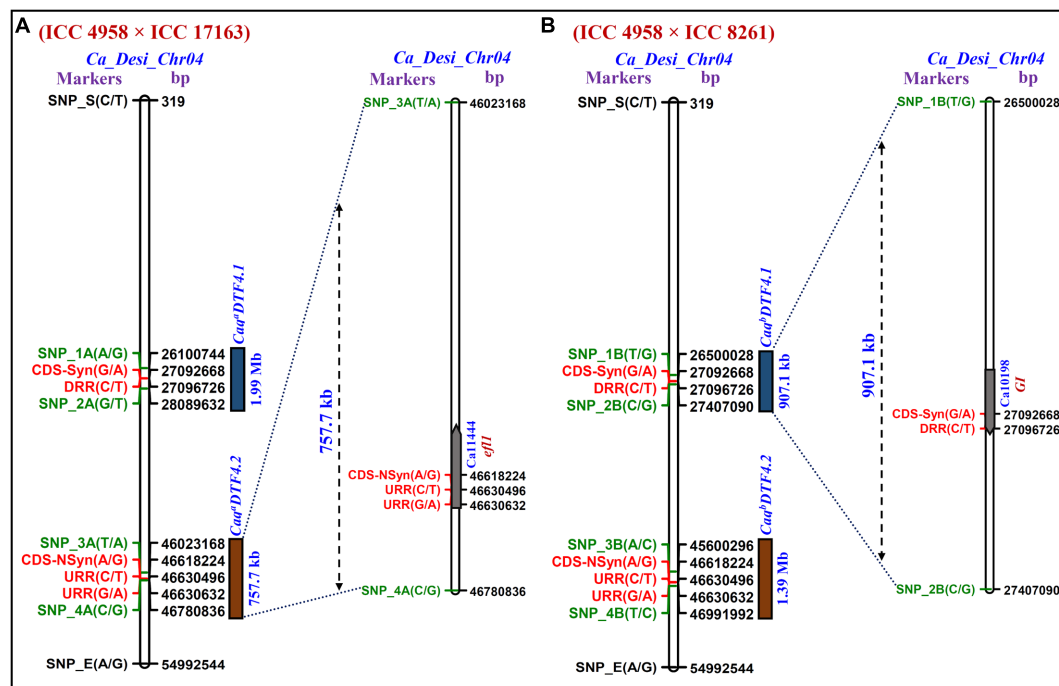


FIGURE 3 | The integration of Δ (SNP-index)-led multiple QTL-seq derived four major flowering time QTLs ($Cac^aDTF4.1$, $Cac^aDTF4.2$, $Cac^bDTF4.1$, and $Cac^bDTF4.2$) in two RIL mapping populations [A: (ICC 4958 \times ICC 17163) and B: (ICC 4958 \times ICC 8261)] scaled-down two longer novel major genomic regions underlying two flowering time QTLs, $Cac^aDTF4.1$ and $Cac^bDTF4.2$ into two smaller 757.7 [between flanking SNP markers: SNP_3A (46023168 bp) to SNP_4A (46780835 bp)] and 907.1 [SNP_1B (26500027 bp) to SNP_2B (27407090 bp)] kb sequence intervals (marked by green font), respectively, on *desi* chromosome 4. Consequently, based on highest Δ (SNP-index) value in these multiple QTL-seq derived DTF QTL intervals, three regulatory and synonymous/non-synonymous coding SNP allelic variants-containing two potential *eff1* and *G1* genes (indicated by red font) regulating flowering time were delineated in chickpea.

TABLE 3 | Gene-derived SNP alleles associated with days to 50% flowering time (DTF) detected by association mapping in chickpea.

SNP IDs*	Kabuli chromosomes	SNP physical positions (bp)	SNPs	Gene accession IDs	Structural annotation	Encoded gene (protein)	Association analysis		
							P	PVE (%)	Traits associated
Ca_LG_446618224	Ca_Kabuli_Chrom4	46618224	A	Ca11444	CDS (Non-synonymous)	<i>eff1</i> (early flowering 1)	1.56×10^{-13}	36	EDTF
		46618224	G				1.43×10^{-12}	38	LDTF
Ca_LG_446630495	Ca_Kabuli_Chrom4	46630495	C	Ca11444	URR	<i>eff1</i> (early flowering 1)	1.07×10^{-12}	49	EDTF
		46630495	T				1.13×10^{-10}	47	LDTF
Ca_LG_446630632	Ca_Kabuli_Chrom4	46630632	G	Ca11444	URR	<i>eff1</i> (early flowering 1)	1.24×10^{-12}	41	EDTF
		46630632	A				1.37×10^{-12}	40	LDTF
Ca_LG_427092669	Ca_Kabuli_Chrom4	27092669	G	Ca10198	CDS (Synonymous)	<i>G1</i> (GIGANTEA)	1.17×10^{-10}	35	EDTF
		27092669	A				1.22×10^{-10}	33	LDTF
Ca_LG_427096726	Ca_Kabuli_Chrom4	27096726	C	Ca10198	DRR	<i>G1</i> (GIGANTEA)	1.30×10^{-11}	43	EDTF
		27096726	T				1.40×10^{-10}	42	LDTF

*Correspond to SNP IDs mentioned in Supplementary Tables S2, S3. EDTF/LDTF, early/late DTF. PVE, phenotypic variation explained.

population phenotyped in field and green house (long- and short-day) conditions at two different geographical locations/years was evident. This infers the complex genetic inheritance pattern of flowering time quantitative trait in chickpea. Therefore,

genetic dissection of this complex quantitative trait employing various genomics-assisted breeding strategies is essential for genetic enhancement and to develop early flowering high seed and pod-yielding stress tolerant cultivars of chickpea during

present scenario of climate change. To accomplish these, our study selectively employed a rapid, cost-efficient and NGS-led high-throughput multiple QTL-seq assay in two inter- and intra-specific RIL mapping population exhibiting a much wider flowering time trait variation including a higher heritability (consistent phenotypic expression) for flowering time in field and green house (long- and short-day) across two diverse geographical locations/years in order to identify major flowering time QTLs in chickpea.

The QTL-seq analysis in an inter-specific RIL mapping population (ICC 4958 \times ICC 17163) detected 1.99 Mb and 757.7 kb two major genomic regions underlying *Caq^aDTF4.1* and *Caq^aDTF4.2* QTLs, respectively, mapped on chromosome 4 governing flowering time in chickpea. Like-wise, 907.1 kb and 1.39 Mb, two major genomic intervals of *Caq^bDTF4.1* and *Caq^bDTF4.2* QTLs, respectively, mapped on chromosome 4 were detected by QTL-seq analysis in an intra-specific RIL mapping population (ICC 4958 \times ICC 8261). These analyses altogether led to identify two consensus major short physical genomic regions of 757.7 kb and 907.1 kb harboring *Caq^aDTF4.2* and *Caq^bDTF4.1* QTLs, respectively, on chromosome 4 of chickpea. The validation of these major DTF QTLs across two diverse intra-/inter-specific chickpea mapping population was apparent implicating the robustness of identified QTLs in regulating flowering time in chickpea. The aforesaid outcomes also inferred the efficacy of multiple QTL-seq assay to narrow-down the longer major DTF QTL intervals detected by QTL-seq in an individual mapping population into shorter major QTL regions in multiple mapping populations of chickpea. This suggests the potential utility of multiple QTL-seq over NGS-based QTL-seq assay and other conventional QTL mapping approaches in high-resolution molecular/fine mapping of major genomic regions harboring QTLs governing diverse agronomic traits including flowering time in chickpea. As per congruent physical positions (bp) on *desi* chromosome 4, two short interval DTF QTLs (*Caq^aDTF4.2* and *Caq^bDTF4.1*) revealed correspondence with the two earlier identified known major flowering time QTLs (*CaqDF4.1* and *CaqDF4.2*) that are identified and mapped on an intra-specific high-density genetic linkage map (ICC 16374 \times ICC 762) of chickpea (Upadhyaya et al., 2015).

The comprehensive multiple QTL-seq analysis at *Caq^aDTF4.2* and *Caq^bDTF4.1* QTL regions detected novel natural allelic variants-containing two strong flowering time-associated *efl1* and *GI* genes with highest Δ (SNP-index) of 1.0 and thereby, considered as the potential candidates for flowering time regulation in chickpea. Two potential candidate genes, *efl1* and *GI* underlying these major QTLs (*Caq^aDTF4.2/CaqDF4.1* and *Caq^bDTF4.1/CaqDF4.2* detected in our present and past studies, respectively) regulating flowering time have been delineated by deploying an integrated genomics-assisted breeding strategy involving candidate gene-based trait association mapping, GWAS, QTL mapping, differential transcript expression profiling and gene-specific molecular haplotyping in chickpea (Upadhyaya et al., 2015). The potential of these identified known flowering development pathway and *FT* gene homologs like *efl1* and *GI* colocalized at the major QTLs in regulating flowering time have

been documented by different traditional QTL mapping studies involving diverse intra- and inter-specific mapping populations of chickpea (Cho et al., 2002; Anbessa et al., 2006; Lichtenzweig et al., 2006; Cobos et al., 2007, 2009; Radhika et al., 2007; Aryamanesh et al., 2010; Hossain et al., 2010; Gowda et al., 2011; Rehman et al., 2011; Hiremath et al., 2012; Vadez et al., 2012; Jamalabadi et al., 2013; Zhang et al., 2013; Varshney et al., 2014). Notably, functional validation and comprehensive molecular characterization of photoperiod-independent *efl1* gene and photoperiod-dependent circadian-clock-related *GI* gene have implicated their potential involvement in regulating flowering time of legumes and *Arabidopsis* (Hecht et al., 2007; Liu et al., 2008; Watanabe et al., 2009, 2011, 2012; Weller et al., 2009; Kang et al., 2010; Laurie et al., 2011; Andres and Coupland, 2012; Kim et al., 2012; Pin and Nilsson, 2012; Song et al., 2013; Yamashino et al., 2013; Zhai et al., 2014; Weller and Martínez, 2015).

Despite of identifying similar flowering time *efl1* and *GI* genes between past and present studies, we were able to discover diverse novel flowering time-regulating non-synonymous and regulatory natural SNP allelic variants (unlike our previous study by Upadhyaya et al., 2015) from the two target *efl1* and *GI* genes that are localized in the two multiple QTL-seq derived major DTF QTL regions (*Caq^aDTF4.2* and *Caq^bDTF4.1*) of chickpea. The detection of altogether different natural allelic variants from two similar *efl1* and *GI* flowering genes localized at two major DTF QTL regions between past and present studies collectively infers the population/cultivar-specific genetic inheritance pattern of complex flowering time quantitative trait in diverse genetic backgrounds of chickpea. The above clues collectively suggest the accuracy, robustness and wider practical applicability of multiple QTL-seq approach for fast genome-wide scanning and mapping of high-resolution major flowering time QTLs as well as delineation of candidate genes and novel natural alleles underlying these major QTLs governing flowering time in chickpea.

The quantitative flowering time trait is primarily governed by complex regulatory networks/pathways involving a diverse array of genes in plant species including legumes (Andres and Coupland, 2012; Song et al., 2013; Weller and Martínez, 2015). The molecular haplotyping of *efl1* and *GI* genes (detected by multiple QTL-seq) among diverse *desi* and *kabuli* cultivated and wild accessions has detected multiple novel natural allelic variants including haplotypes in these flowering time genes exhibiting varied potential characteristics for flowering time trait regulation and evolutionary pattern in domesticated chickpea (Upadhyaya et al., 2015). Therefore, novel functionally relevant potential molecular signatures (SNP markers, genes, QTLs and natural allelic variants) governing flowering time delineated by us employing a NGS-based high-throughput multiple QTL-seq strategy can be useful for fast genetic dissection of complex flowering time quantitative trait and eventually genomics-assisted crop improvement to develop early flowering varieties of chickpea with limited resource expenses.

Preliminary efforts have been made to understand the natural/functional allelic diversity-based domestication pattern

of two flowering (*efl1* and *GI*) genes localized at two major DTF QTL regions among *Cicer* cultivated (*desi* and *kabuli*) and wild gene pools. This exhibited a significant reduction of natural/functional allelic diversity in cultivated *desi* and *kabuli* accessions from diverse geographical regions of the world as compared to annual and perennial wild accessions belonging to primary, secondary and tertiary gene pools of chickpea. However, this observed background allelic diversity reduction in cultivated than that of wild chickpea was much stronger especially at major DTF QTL regions where *efl1* and *GI* flowering genes were mapped. This implicates that these genes were targeted by artificial selection which was further evident from their non-neutral evolution during chickpea domestication based on significant variation of Tajima's D between cultivated (−1.12) and wild (0.26) accessions, respectively. Interestingly, all these natural allelic variants-containing two potential genes localized within flowering time major QTL intervals have been commonly identified and mapped on multiple independent chickpea mapping populations by earlier and our present studies (Upadhyaya et al., 2015). These outcomes clearly reflect the extensive contribution of four sequential evolutionary bottlenecks including vernalization and strong artificial and/or natural selection pressure on these flowering time-associated natural allelic variants of two gene loci (*efl1* and *GI*) during chickpea domestication leading toward reduction of genetic diversity in cultivated chickpea as compared to that of wild *Cicer* gene pool (Lev-Yadun et al., 2000; Abbo et al., 2003; Berger et al., 2005; Burger et al., 2008; Toker, 2009; Meyer et al., 2012; Jain et al., 2013; Kujur et al., 2013; Varshney et al., 2013; Saxena et al., 2014a,b).

The most crucial evolutionary bottleneck, vernalization, is a vital key module of flowering time during chickpea domestication culminating into existence of currently cultivated vernalization insensitive *desi* and *kabuli* cultivars specifically from the vernalization sensitive wild ancestor *C. reticulatum* (Abbo et al., 2003, 2014; Berger et al., 2005; Burger et al., 2008; Toker, 2009). The major domestication bottlenecks integrated with artificial selection including modern breeding efforts have been constantly practiced during the chickpea genetic improvement program for developing its early flowering cultivars of high seed and pod yield. These findings collectively infer that the natural allelic variation-based functional diversity scanned in the genes might be associated with flowering time trait evolution with regard to differential domestication-led bottlenecks including vernalization response in *desi* and *kabuli* cultivated and wild chickpea during domestication. Henceforth, flowering time represents a vital component of domestication trait selected during breeding and genetics of chickpea. Moreover, the major impact of long- and short-photoperiods which are the major environmental cues for determining the flowering time including time of flower initiation and/or first flower

appearance especially in photoperiod-sensitive as compared to photoperiod-insensitive chickpea accessions is well documented (Daba et al., 2016). In the present study, a significant interactions between long- and short-day photoperiods and DTF trait variation observed in individuals of two RIL mapping populations across 2 years was apparent. In spite of this concern, we were able to identify functionally relevant non-synonymous/synonymous coding and regulatory SNP allelic variants from two flowering genes (*efl1* and *GI*) localized at two major DTF QTL regions by using the long- and short-day phenotyping data of both RIL populations separately in multiple QTL-seq assay. This further infers the efficacy of strategy (multiple QTL-seq) implemented in our study to detect potential molecular signatures regulating flowering time in chickpea. It is, therefore, essential to perform a comprehensive analysis using all natural/functional allelic variants discovered and potential locus targeted by natural and/or artificial selection in two flowering genes (*efl1* and *GI*) to delve deeper into the complex flowering time trait evolution and domestication in chickpea. This will be useful to understand the molecular mechanism influencing fixation of such complex flowering time quantitative trait in domesticated cultivars that are adapted to multiple agro-ecological regions of the world and further pave the way for genetic enhancement to develop early flowering high seed/pod-yielding varieties of chickpea amidst current climate change scenario.

AUTHOR CONTRIBUTIONS

RS, AD, and UB conducted all experiments and drafted the manuscript. HU, RK, PS, ST, and CB helped in development, advancement and phenotyping of mapping populations. AT and SP conceived and designed the study, guided data analysis and interpretation, participated in drafting and correcting the manuscript critically and gave the final approval of the version to be published. All authors have read and approved the final manuscript.

ACKNOWLEDGMENT

The financial support by the Department of Biotechnology (DBT), Government of India to NIPGR is acknowledged.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.01105/full#supplementary-material>. For Tables S2 and S3, see: <https://figshare.com/s/f1848022c4f3c620f1d2>.

REFERENCES

- Abbo, S., Berger, J., and Turner, N. C. (2003). Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.* 30, 1081–1087. doi: 10.1071/FP03084
- Abbo, S., Pinhasi van-Oss, R., Gopher, A., Saranga, Y., Ofner, I., and Peleg, Z. (2014). Plant domestication versus crop evolution: a conceptual framework for cereals and grain legumes. *Trends Plant Sci.* 19, 351–360. doi: 10.1016/j.tplants.2013.12.002
- Anbessa, Y., Warkentin, T., Vandenberg, A., and Ball, R. (2006). Inheritance of time to flowering in chickpea in a short-season temperate environments. *J. Hered.* 97, 55–61. doi: 10.1093/jhered/esj009
- Andres, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* 13, 627–639. doi: 10.1038/nrg3291
- Aryamanesh, N., Nelson, M. N., Yan, G., Clarke, H. J., and Siddique, K. H. M. (2010). Mapping a major gene for growth habit and QTLs for *Ascochyta* blight resistance and flowering time in a population between chickpea and *Cicer reticulatum*. *Euphytica* 173, 307–319. doi: 10.1007/s10681-009-0086-2
- Bajaj, D., Das, S., Upadhyaya, H. D., Ranjan, R., Badoni, S., Kumar, V., et al. (2015a). A genome-wide combinatorial strategy dissects complex genetic architecture of seed coat color in chickpea. *Front. Plant Sci.* 6:979. doi: 10.3389/fpls.2015.00979
- Bajaj, D., Upadhyaya, H. D., Khan, Y., Das, S., Badoni, S., Shree, T., et al. (2015b). A combinatorial approach of comprehensive QTL-based comparative genome mapping and transcript profiling identified a seed weight-regulating candidate gene in chickpea. *Sci. Rep.* 5:9264. doi: 10.1038/srep09264
- Berger, J. D., Buck, R., Henzell, J. M., and Turner, N. C. (2005). Evolution in the genus *Cicer* vernalisation response and low temperature pod set in chickpea (*C. arietinum* L.) and its annual wild relatives. *Aust. J. Agric. Res.* 56, 1191–1200. doi: 10.1071/AR05089
- Burger, J. C., Champan, M. A., and Burke, J. M. (2008). Molecular insights into the evolution of crop plants. *Am. J. Bot.* 95, 113–122. doi: 10.3732/ajb.95.2.113
- Cho, S., Kumar, J., Shultz, J. F., Anupama, K., Tefera, F., and Muehlbauer, F. J. (2002). Mapping genes for double podding and other morphological traits in chickpea. *Euphytica* 125, 285–292. doi: 10.1023/A:1020872009306
- Cobos, M. J., Rubio, J., Fernández-Romero, M. D., Garza, R., Moreno, M. T., Millán, T., et al. (2007). Genetic analysis of seed size, yield and days to flowering in a chickpea recombinant inbred line population derived from a *kabuli* x *desi* cross. *Ann. Appl. Biol.* 151, 33–42. doi: 10.1111/j.1744-7348.2007.00152.x
- Cobos, M. J., Winter, P., Kharrat, M., Cubero, J. I., Gil, J., and Millán, T. (2009). Genetic analysis of agronomic traits in a wide cross of chickpea. *Field Crops Res.* 111, 130–136. doi: 10.1016/j.fcr.2008.11.006
- Daba, K., Warkentin, T. D., Bueckert, R., Todd, C. D., and Tar'an, B. (2016). Determination of photoperiod-sensitive phase in chickpea (*Cicer arietinum* L.). *Front. Plant Sci.* 7:478. doi: 10.3389/fpls.2016.00478
- Das, S., Singh, M., Srivastava, R., Bajaj, D., Saxena, M. S., Rana, J. C., et al. (2016). mQTL-seq delineates functionally relevant candidate gene harbouring a major QTL regulating pod number in chickpea. *DNA Res.* 23, 53–65. doi: 10.1093/dnares/dsv036
- Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Laxmi, et al. (2015a). Deploying QTL-seq for rapid delineation of a potential candidate gene underlying major trait-associated QTL in chickpea. *DNA Res.* 22, 193–203. doi: 10.1093/dnares/dsv004
- Das, S., Upadhyaya, H. D., Srivastava, R., Bajaj, D., Gowda, C. L. L., Sharma, S., et al. (2015b). Genome-wide insertion-deletion (InDel) marker discovery and genotyping for genomics-assisted breeding applications in chickpea. *DNA Res.* 22, 377–386. doi: 10.1093/dnares/dsv020
- Gaur, P. M., Samineni, S., Tripathi, S., Varshney, R. K., and Gowda, C. L. L. (2014). Allelic relationships of flowering time genes in chickpea. *Euphytica* 203, 295–308. doi: 10.1007/s10681-014-1261-7
- Gowda, C. L. L., Upadhyaya, H. D., Dronavalli, N., and Singh, S. (2011). Identification of large-seeded high-yielding stable *kabuli* chickpea germplasm lines for use in crop improvement. *Crop Sci.* 5, 198–209. doi: 10.2135/cropsci2010.01.0078
- Hecht, V., Knowles, C. L., Vander Schoor, J. K., Liew, L. C., Jones, S. E., Lambert, M. J., et al. (2007). Pea *LATE BLOOMER1* is a *GIGANTEA* ortholog with roles in photoperiodic flowering, deetiolation, and transcriptional regulation of circadian clock gene homologs. *Plant Physiol.* 144, 648–661. doi: 10.1104/pp.107.096818
- Hegde, V. S. (2010). Genetics of flowering time in chickpea in a semi-arid environment. *Plant Breed.* 129, 683–687. doi: 10.1111/j.1439-0523.2009.01748.x
- Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10, 716–732. doi: 10.1111/j.1467-7652.2012.00710.x
- Hossain, S., Ford, R., McNeil, D. L., Pittock, C., and Pannozzo, J. F. (2010). Development of a selection tool for seed shape and QTL analysis of seed shape with other morphological traits for selective breeding in chickpea (*Cicer arietinum* L.). *Aust. J. Crop Sci.* 4, 278–288.
- Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tpj.12173
- Jamalabadi, J. G., Saidi, A., Karami, E., Kharkesh, M., and Talebi, R. (2013). Molecular mapping and characterization of genes governing time to flowering, seed weight, and plant height in an intraspecific genetic linkage map of chickpea (*Cicer arietinum*). *Biochem. Genet.* 51, 387–397. doi: 10.1007/s10528-013-9571-3
- Kang, J., Hwang, J. U., Lee, M., Kim, Y. Y., Assmann, S. M., Martinoia, E., et al. (2010). PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proc. Natl. Acad. Sci. U.S.A.* 107, 2355–2360. doi: 10.1073/pnas.0909222107
- Kashiwagi, J., Krishnamurthy, L., Gaur, P. M., Upadhyaya, H. D., Varshney, R. K., and Tobita, S. (2013). Traits of relevance to improve yield under terminal drought stress in chickpea (*C. arietinum* L.). *Field Crops Res.* 145, 88–95. doi: 10.1016/j.fcr.2013.02.011
- Kim, M. Y., Shin, J. H., Kang, Y. J., Shim, S. R., and Lee, S. H. (2012). Divergence of flowering genes in soybean. *J. Biosci.* 37, 857–870. doi: 10.1007/s12038-012-9252-0
- Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–374. doi: 10.1093/dnares/dst015
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015a). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Front. Plant Sci.* 6:162. doi: 10.3389/fpls.2015.00162
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015b). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5:11166. doi: 10.1038/srep11166
- Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M. S., et al. (2015c). Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci. Rep.* 5:9468. doi: 10.1038/srep09468
- Kumar, A., Choudhary, A. K., Solanki, R. K., and Pratap, A. (2011). Towards marker-assisted selection in pulses: a review. *Plant Breed.* 130, 297–313. doi: 10.1007/s00299-017-2127-y
- Laurie, R. E., Diwadkar, P., Jaudal, M., Zhang, L., Hecht, V., Wen, J., et al. (2011). The *Medicago truncatula* flowering locus T homologue, *MtFTa1*, is a key regulator of flowering time. *Plant Physiol.* 156, 2207–2224. doi: 10.1104/pp.111.180182
- Lev-Yadun, S., Gopher, A., and Abbo, S. (2000). The cradle of agriculture. *Science* 288, 1602–1603. doi: 10.1126/science.288.5471.1602
- Lichtenzweig, J., Bonfil, D. J., Zhang, H. B., Shtienberg, D., and Abbo, S. (2006). Mapping quantitative trait loci in chickpea associated with time to flowering and resistance to *Didymella rabiei* the causal agent of *Ascochyta* blight. *Theor. Appl. Genet.* 113, 1357–1369. doi: 10.1007/s00122-006-0390-3
- Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., and Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180, 995–1007. doi: 10.1534/genetics.108.092742
- Lu, H., Lin, T., Klein, J., Wang, S., Qi, J., Zhou, Q., et al. (2014). QTL-seq identifies an early flowering QTL located near *FLOWERING LOCUS T* in cucumber. *Theor. Appl. Genet.* 127, 1491–1499. doi: 10.1007/s00122-014-2313-z

- Meyer, R. S., DuVal, A. E., and Jensen, H. R. (2012). Patterns and processes in crop domestication: an historical review and quantitative analysis of 203 global food crops. *New Phytol.* 196, 29–48. doi: 10.1111/j.1469-8137.2012.04253.x
- Parween, S., Nawaz, K., Roy, R., Pole, A. K., Venkata Suresh, B., Misra, G., et al. (2015). An advanced draft genome assembly of a *desi* type chickpea (*Cicer arietinum* L.). *Sci. Rep.* 5:12806. doi: 10.1038/srep12806
- Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS ONE* 7:e30619. doi: 10.1371/journal.pone.0030619
- Pierre, J. B., Bogard, M., Herrmann, D., Huyghe, C., and Julier, B. (2011). A CONSTANS-like gene candidate that could explain most of the genetic variation for flowering date in *Medicago truncatula*. *Mol. Breed.* 28, 25–35. doi: 10.1007/s11032-010-9457-6
- Pierre, J. B., Huguet, T., Barre, P., Huyghe, C., and Julier, B. (2008). Detection of QTLs for flowering date in three mapping populations of the model legume species *Medicago truncatula*. *Theor. Appl. Genet.* 117, 609–620. doi: 10.1007/s00122-008-0805-4
- Pin, P. A., and Nilsson, O. (2012). The multifaceted roles of *FLOWERING LOCUS T* in plant development. *Plant Cell Environ.* 35, 1742–1755. doi: 10.1111/j.1365-3040.2012.02558.x
- Radhika, P., Gowda, S. J. M., Kadoo, N. Y., Mhase, L. B., Jamadagni, B. M., Sainani, M. N., et al. (2007). Development of an integrated intraspecific map of chickpea (*Cicer arietinum* L.) using two recombinant inbred line populations. *Theor. Appl. Genet.* 115, 209–216. doi: 10.1007/s00122-007-0556-7
- Rehman, A. U., Malhotra, R. S., Bett, K., Tar'an, B., Bueckert, R., and Warkentin, T. D. (2011). Mapping QTL associated with traits affecting grain yield in chickpea (*Cicer arietinum* L.) under terminal drought stress. *Crop Sci.* 51, 450–463. doi: 10.2135/cropsci2010.03.0129
- Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014a). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031
- Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014b). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484
- Singh, V. K., Garg, R., and Jain, M. (2013). A global view of transcriptome dynamics during flower development in chickpea by deep sequencing. *Plant Biotechnol. J.* 11, 691–701. doi: 10.1111/pbi.12059
- Singh, V. K., Khan, A. W., Jaganathan, D., Thudi, M., Roorkiwal, M., Takagi, H., et al. (2016). QTL-seq for rapid identification of candidate genes for 100-seed weight and root/total plant dry weight ratio under rainfed conditions in chickpea. *Plant Biotechnol. J.* 14, 2110–2119. doi: 10.1111/pbi.12567
- Song, Y. H., Ito, S., and Imaizumi, T. (2013). Flowering time regulation: photoperiod- and temperature-sensing in leaves. *Trends Plant Sci.* 18, 575–583. doi: 10.1016/j.tplants.2013.05.003
- Stephens, A., Lombardi, M., Cogan, N. O. I., Foster, J. W., Hobson, K., Materne, M., et al. (2014). Genetic marker discovery, intra-specific linkage map construction and quantitative trait locus analysis of *Ascochyta* blight resistance in chickpea (*Cicer arietinum* L.). *Mol. Breed.* 33, 297–313. doi: 10.1007/s11032-013-9950-9
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoaka, C., et al. (2013). QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. *Plant J.* 74, 174–183. doi: 10.1111/tpj.12105
- Toker, C. (2009). A note on the evolution of kabuli chickpeas as shown by induced mutations in *Cicer reticulatum* Ladizinsky. *Genet. Resour. Crop Evol.* 56, 7–12. doi: 10.1007/s10722-008-9336-8
- Upadhyaya, H. D., Bajaj, D., Das, S., Saxena, M. S., Badoni, S., Kumar, V., et al. (2015). A genome-scale integrated approach aids in genetic dissection of complex flowering time trait in chickpea. *Plant Mol. Biol.* 89, 403–420. doi: 10.1007/s11103-015-0377-z
- Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* 41, 206–210. doi: 10.2135/cropsci2001.411206x
- Upadhyaya, H. D., Dwivedi, S. L., Baum, M., Varshney, R. K., Udupa, S. M., and Gowda, C. L. L. (2008). Genetic structure, diversity and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 8:106. doi: 10.1186/1471-2229-8-106
- Vadez, V., Krishnamurthy, L., Thudi, M., Anuradha, C., Colmer, T. D., Turner, N. C., et al. (2012). Assessment of ICCV2 × JG62 chickpea progenies shows sensitivity of reproduction to salt stress and reveals QTLs for seed yield and yield components. *Mol. Breed.* 30, 9–21. doi: 10.1007/s11032-011-9594-6
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., and Krishnamurthy, L. (2014). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6
- Watanabe, S., Harada, K., and Abe, J. (2012). Genetic and molecular bases of photoperiod responses of flowering in soybean. *Breed. Sci.* 61, 531–543. doi: 10.1270/jsbbs.61.531
- Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus *E3*. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772
- Watanabe, S., Xia, Z., Hideshima, R., Tsubokura, Y., Sato, S., and Yamanaka, N. (2011). A map-based cloning strategy employing a residual heterozygous line reveals that the *GIGANTEA* gene is involved in soybean maturity and flowering. *Genetics* 188, 395–407. doi: 10.1534/genetics.110.125062
- Weller, J. L., Hecht, V., Liew, L. C., Sussmilch, F. C., Wenden, B., Knowles, C. L., et al. (2009). Update on the genetic control of flowering in garden pea. *J. Exp. Bot.* 60, 2493–2499. doi: 10.1093/jxb/erp120
- Weller, J. L., and Martínez, R. O. (2015). Genetic control of flowering time in legumes. *Front. Plant Sci.* 6:207. doi: 10.3389/fpls.2015.00207
- Yamashino, T., Yamawaki, S., Hagui, E., Ueoka-Nakanishi, H., Nakamichi, N., Ito, S., et al. (2013). Clock-controlled and *FLOWERING LOCUS T* (FT)-dependent photoperiodic pathway in *Lotus japonicus* I: verification of the flowering-associated function of an FT homolog. *Biosci. Biotechnol. Biochem.* 77, 747–753. doi: 10.1271/bbb.120871
- Zaman-Allah, M., Jenkinson, D. M., and Vadez, V. (2011). Chickpea genotypes contrasting for seed yield under terminal drought stress in the field differ for traits related to the control of water use. *Funct. Plant Biol.* 38, 270–281. doi: 10.1071/FP10244
- Zhai, H., Lu, S., Liang, S., Wu, H., Zhang, X., Liu, B., et al. (2014). *GmFT4*, a homolog of *FLOWERING LOCUS T*, is positively regulated by *E1* and functions as a flowering repressor in soybean. *PLoS ONE* 9:e89030. doi: 10.1371/journal.pone.0089030
- Zhang, D., Cheng, H., Hu, Z., Wang, H., Kan, G., Liu, C., et al. (2013). Fine mapping of a major flowering time QTL on soybean chromosome 6 combining linkage and association analysis. *Euphytica* 191, 23–33. doi: 10.1007/s10681-012-0840-8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Srivastava, Upadhyaya, Kumar, Daware, Basu, Shimray, Tripathi, Bharadwaj, Tyagi and Parida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Fine-Mapping the Branching Habit Trait in Cultivated Peanut by Combining Bulk Segregant Analysis and High-Throughput Sequencing

Galya Kayam, Yael Brand, Adi Faigenboim-Doron, Abhinandan Patil, Ilan Hedvat and Ran Hovav*

Department of Field Crops, Plant Science Institute, Agricultural Research Organization, Bet-Dagan, Israel

OPEN ACCESS

Edited by:

Maria Carlota Vaz Patto,
Instituto de Tecnologia Química e
Biológica and Universidade Nova de
Lisboa, Portugal

Reviewed by:

Daniela Marone,
Centre of Cereal
Research-CREA-CER-Foggia, Italy
Teresa Millan,
Universidad de Cordoba, Spain
Daniel Fonceka,
Agricultural Research Centre For
International Development, France

*Correspondence:

Ran Hovav
ranh@agri.gov.il

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 28 December 2016

Accepted: 16 March 2017

Published: 04 April 2017

Citation:

Kayam G, Brand Y,
Faigenboim-Doron A, Patil A, Hedvat I
and Hovav R (2017) Fine-Mapping the
Branching Habit Trait in Cultivated
Peanut by Combining Bulk
Segregant Analysis and
High-Throughput Sequencing.
Front. Plant Sci. 8:467.
doi: 10.3389/fpls.2017.00467

The growth habit of lateral shoots (also termed “branching habit”) is an important descriptive and agronomic character of peanut. Yet, both the inheritance of branching habit and the genetic mechanism that controls it in this crop remain unclear. In addition, the low degree of polymorphism among cultivated peanut varieties hinders fine-mapping of this and other traits in non-homozygous genetic structures. Here, we combined high-throughput sequencing with a well-defined genetic system to study these issues in peanut. Initially, segregating F_2 populations derived from a reciprocal cross between very closely related Virginia-type peanut cultivars with spreading and bunch growth habits were examined. The spreading/bunch trait was shown to be controlled by a single gene with no cytoplasmic effect. That gene was named *Bunch1* and was significantly correlated with pod yield per plant, time to maturation and the ratio of “dead-end” pods. Subsequently, bulk segregant analysis was performed on 52 completely bunch, and 47 completely spreading F_3 families. In order to facilitate the process of SNP detection and candidate-gene analysis, the transcriptome was used instead of genomic DNA. Young leaves were sampled and bulked. Reads from Illumina sequencing were aligned against the peanut reference transcriptome and the diploid genomes. Inter-varietal SNPs were detected, scored and quality-filtered. Thirty-four candidate SNPs were found to have a bulk frequency ratio value >10 and 6 of those SNPs were found to be located in the genomic region of linkage group B5. Three best hits from that over-represented region were further analyzed in the segregating population. The trait locus was found to be located in a ~ 1.1 Mbp segment between markers M875 (B5:145,553,897; 1.9 cM) and M255 (B5:146,649,943; 2.25 cM). The method was validated using a population of recombinant inbred lines of the same cross and a new DNA SNP-array. This study demonstrates the relatively straight-forward utilization of bulk segregant analysis for trait fine-mapping in the low polymeric and heterozygous germplasm of cultivated peanut and provides a baseline for candidate gene discovery and map-based cloning of *Bunch1*.

Keywords: peanut, branching habit, bulk segregant analysis, fine mapping

INTRODUCTION

Peanut (*Arachis hypogaea* L.) is an economically significant crop grown throughout the world. It is the second-most important cultivated grain legume and the fourth largest edible oilseed crop (Faostat, 1998). It is an unusual legume plant in that its flowers are borne aboveground, but the fruits develop underground. The plant is an indeterminate, annual herbaceous bush that is 15–70 cm tall and is comprised of an erect main shoot and a number of lateral shoots (branches) that begin at the base of the plant.

The growth angle of the lateral shoots (commonly referred to as “growth habit” or “branching habit”) is one of the most important descriptive characteristics of peanut (Pittman, 1995). The wild polyploid peanut species *A. monticola* usually has a spreading phenotype, in which lateral branches strike or partially strike the ground. In domesticated peanut, four different, and easily distinguishable categories of branching habit are known: prostrate, spreading, bunch, and erect (Table 1). The description of a new peanut variety, especially one that is of the Virginia marketing-type, will almost always begin with the definition of its growth habit (e.g., bunch or spreading). In addition to assisting breeders and other researchers in identifying accessions for specific traits, branching habit has a great impact on peanut physiology, productivity, and crop management. Since fruiting in peanut occurs underground, the distance between the flowering buds and the ground is an important factor. Pegs of bunch/erect plants that do not reach the ground will not produce pods on time. However, the pods of bunch/erect plants develop at the same time, promoting early maturation. The growth habit of peanut affects the implementation of agrotechnology such as mechanical cultivation and disease management (Butzler et al., 1998).





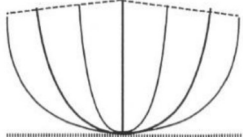

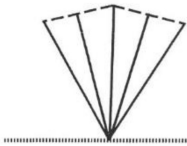

Despite the agronomic importance of the growth habit of peanut, both the inheritance of this trait and the genetic mechanism that controls branching habit in this crop are not clear. The trait was studied in detail during the 1960s and 1970s and, during that period, two distinct phenotypic groups were usually considered: the runner growth habit wherein the side branches are prostrate, always growing peripherally from the main axis, trailing on the ground except for the tips, which may be somewhat ascending and the bunch growth habit, in which the laterals are also erect or ascending. Initially, a two-gene model for the control of growth habit was suggested, with the runner habit dominant to the erect habit (Hull, 1933; Patel et al., 1936; Coffelt, 1974), but investigators had difficulty classifying the intermediate and/or abnormal growth habits of F₂ progeny of crosses between plants exhibiting different growth habits. A fundamental set of experiments conducted by Ashri (1964, 1968) indicated the existence of a genic-cytoplasmic interaction that controls growth habit in peanut. In a project involving a series of reciprocal crosses, differences in growth habit were recorded. A few more nuclear and cytoplasmic genes were later identified by the same group and those researchers concluded that cytoplasmic inheritance has a major effect on the branching habit of peanut (Ashri, 1975; Ashri and Levi, 1975). In more recent studies, the branching habit trait was genetically characterized and mapped by using inter-specific crossing system with an amphidiploid

species (Fonceka et al., 2012a,b). The branching habit trait was phenotyped quantitatively by using a continuous scale from 1 (procumbent) to 6 (erect). The trait showed a wide range of morphologies, ranging from completely prostrate to totally erect, and several QTLs were found to control the trait, with the most significant located on linkage groups a07, b05, a10, and b10.

Like many polyploid species, cultivated peanut has experienced a genetic bottleneck, which, together with the effects of domestication, has greatly narrowed its genetic diversity and limited DNA polymorphism among subsequently derived *Arachis* lines (Kochert et al., 1991; Moretzsohn et al., 2013). As a result, peanut has a low degree of polymorphism among cultivated varieties. This limited polymorphism has hindered the development of molecular and genomic tools for use in domesticated peanut. With the introduction of the genome sequences of peanut ancestors *Arachis duranensis* and *Arachis ipaensis* (Bertioli et al., 2016), peanut is now “the orphan legume genome whose time has come” (Ozias-Akins, 2013). These genome sequences provide the resources necessary to move peanut genomics to the next level, facilitating the development of SNP-based marker technologies. In the past, the most widely used molecular markers were simple-sequence repeats (SSRs). Despite their widespread use on the intra-species cultivated level (e.g., Selvaraj et al., 2009), the utility of SSR studies of peanut is limited by their apparent low frequency of across the genome and the relatively low-throughput method of analysis. The use of high-throughput markers like SNPs is necessary for efficient use of genomic data for marker-assisted selection, quantitative trait locus mapping and genomic selection. Recently, several platforms have been developed to facilitate the use of SNP markers for gene-mapping in peanut, including genotyping by sequencing (Zhou et al., 2014) and genome resequencing-based SNP arrays (Clevenger et al., 2016; Pandey et al., 2017). However, despite these advances, these platforms are usually highly efficient only for homozygous populations, like recombinant inbred lines (RILs) and introgression lines (ILs), which are relatively tedious and expensive to construct in peanut. These methods are usually less effective for trait-mapping in heterozygous genetic populations (e.g., F₂ and F₃ generations) due to the allopolyploid nature of the peanut genome. This is particularly true in cases of genetic populations that are based upon a cross between closely related parental genotypes, which are occasionally needed for better genetic dissection of specific traits.

In this study, we used a well-defined genetic system to further investigate the genetic nature of the branching habit trait of peanut. Initially, segregating F₂ populations derived from a reciprocal cross between very closely-related Virginia marketing-type cultivars were analyzed. Against this particular background, the spreading/bunch trait, a well-known characteristic of the Virginia varieties, was found to be controlled by a single gene with no cytoplasmic effect. Subsequently, a combination of bulked segregant analysis and deep sequencing was developed to facilitate the SNP detection process and the fine-mapping of this gene. The processes were validated using a RIL population derived from the same cross and a new SNP array (Pandey et al., 2017). The relatively straight-forward utilization of this

TABLE 1 | The four main types of growth habit in peanut (from Pittman, 1995).

Phenotype	Illustration	Description	Example
Prostrate		Branches that strike the ground and a conspicuous main stem	
Spreading		Branches partially on the ground, with their tips curved upward	
Bunch		Semi-erect, branches curved upward, beginning at the base; main stem slightly taller than the others	
Erect		Branches grow straight up from the base and are generally 45° or less from vertical (<i>TAC1</i> style)	

The two growth habits represented in this study are marked in red.

technique in ultra-low polymorphic and highly heterozygous peanut germplasm is demonstrated.

MATERIALS AND METHODS

Plant Material and Data Collection

Segregating F₂ and F₃ populations derived from a reciprocal cross between very closely-related Virginia-type peanut cultivars were studied. The parental lines were cv. “Hanoch,” a late-maturing spreading-type cultivar, and cv. “Harari,” a medium-maturing bunch-type cultivar. The two parental lines share substantial genetic background, since cv. Harari was developed from an initial cross between cv. Hanoch and cv. Shulamit and an additional back-cross of cv. Hanoch with cv. Hillah (the outcome of Hanoch × Shulamit). In 2013, 314, and 252 F₂ individuals from reciprocal Hanoch × Harari and Harari × Hanoch crosses, respectively, were grown under field conditions. The plot consisted of two rows, 75 cm apart, with 40 cm spacing between plants within each row. The experimental plots were sown alongside commercial plots under full irrigation. All agricultural practices were carried out according to local growing protocols as described previously (Gupta et al., 2014).

Growth habit was recorded at 80 days after sowing. At the end of the season, pods were harvested on an individual-plant basis. For each sample, the total pod yield, net pod yield (where immature and unhealthy pods are excluded), number of pods, total seed weight, “dead-end” ratio (relative number of pods with the remote seed aborted) and seed ratio (net seed weight per plant divided by the net pod weight per plant) traits were recorded as well. From each population, approximately 150 F₃ families were grown in the subsequent season, with 16 seeds from each F₂ individual sown. Branching habit was recorded at

the family level (spreading/bunch/segregating) at 80 days after sowing. Plant maturity index was determined based on three random plants in the homozygous F₃ families, as the number of fully matured pods out of the total number of pods at 140 days after sowing. To validate the bulk segregant analysis, 94 RILs (F₆:F₈), which originated from the same Hanoch × Harari cross by single seed descent, were analyzed. In 2016, 16 randomly-selected plants from each RIL were grown under field conditions and the growth habit of each plant was recorded 70 days after sowing.

RNA Isolation, Preparation of Libraries, and High-Throughput Sequencing

Bulked segregant analysis was performed on the F₃ families that were found to be homozygous for the spreading or bunch growth habit. In total, 52 completely bunch and 47 completely spreading families were sampled. Young leaves were collected from all 16 individuals in each family. In each phenotypic group (spreading/bunch), all tissues from the families were bulked for the RNA extraction. Working on the RNA level was preferable to working on the DNA genomic level due to the large and relatively complex peanut genome and also facilitated the detection of candidate genes. Samples were taken of each of the ground tissues (400 mg each) and were used for RNA extraction using the hot-borate method, as described by Brand and Hovav (2010). The total RNA was used to prepare two RNA-Seq libraries, using TruSeq RNA Sample Preparation Kit v2 (Illumina) following the manufacturer’s protocol as described previously (Gupta et al., 2016). Libraries were validated using DNA Screen Tape D1000 and the TapeStation 2200 (Agilent). RNA-Seq libraries were sequenced using an Illumina HiSeq™2000 (single lane) at the sequencing center at the Technion in Haifa, Israel.

Data analyses followed the general guidelines for bulk segregant analysis using next-generation sequencing (Magwene et al., 2011) and the specific guidelines for polyploids (Trick et al., 2012), with several modifications. Raw reads were subjected to a cleaning procedure using the FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.htm) including: (1) trimming read-end nucleotides with quality scores <30 using `fastq_quality_trimmer` and (2) removing reads with less than 70% base pairs with quality score ≤ 30 using `fastq_quality_filter`. The sequences were mapped against the 4X tetraploid peanut transcript assembly reference (<http://www.peanutbase.org/>) and against two *Arachis* diploid genomes (*A. duranensis* and *A. ipaensis*; Bertoli et al., 2016; peanutbase.org) using Bowtie2 aligner (Langmead and Salzberg, 2012). The genome Analysis Toolkit (GATK) Unified Genotyper software version 2.5.2 (McKenna et al., 2010; DePristo et al., 2011) was used for the detection of SNPs. A custom Perl script was used to derive the symmetric difference of the two SNP sets. Polymorphisms between homologous genomes generate the same doubled code and should be common to both SNP sets. Yet, differences in the SNPs between cv. Hanoch and cv. Harari (varietal-specific SNPs) should generate doubled code for only one bulk and, therefore, be unique to the corresponding SNP set. In this manner, ~13,000 varietal-specific SNPs were retrieved between the two bulks. These SNPs were further filtered according to the number of reads for each SNP > 50, GATK quality value >100 and BFR >3. Also, genes with SNP densities higher than 5 SNPs/kb were eliminated to avoid possible paralogue SNPs.

Validation of the SNP Markers and the Bulk Segregant Analysis

For further validation of the SNPs, DNA was collected from the parental lines and 20 F_3 progeny of the cross cv. Hanoch \times cv. Harari using a DNA Easy kit (SIGMA Aldrich). The same leaves that were used for the RNA study were used for DNA extraction, but the DNA analysis was conducted on a single-plant basis instead of with bulk samples. To validate the three best SNP markers, the following primers were used: M35: F-TCTC TCTCTCTCACAGTCAC; R-CTTGCCGGCAAATAGAGCAT. M255: F--CAGATATGCAAGGCCTAACT; R-TGCCAGAGCA AGGAACATGT. M875: F-CCATCTGCAGTGAGAGTCAA; R-GTGATTCCTGCGTTCAAGTC. These primers were also used for further mapping of the trait in 182 F_4 individuals derived from one F_3 family segregating for the branching habit trait.

The fine-mapping of the branching habit gene carried out using the bulk-segregant approach was further validated by a custom Affymetrix Axiom SNP array (Pandey et al., 2017). For that analysis, DNA was collected from the two parental lines and each of 94 Recombinant Inbred Lines (RILs) derived from the same cv. Hanoch \times cv. Harari cross. Young leaves were collected from 12 random plants from each RIL and DNA was extracted with a specific kit (GenElute™; Sigma). DNA was quantified by Qubit (Invitrogen LTD) and diluted to 30 ng/uL according to the Affymetrix guidelines (<http://www.affymetrix.com>). The chip array calls were subjected to cluster-quality filtering, carried out

according to Affymetrix guidelines, and additional filtering to select only SNPs that were polymorphic between the parental lines and segregated in a 1:1 ratio in the RILs.

RESULTS

Branching Habit Is Controlled by a Single Gene

The segregation patterns of the branching habit trait in the F_1 , F_2 , and F_3 generations of the “Hanoch” \times “Harari” cross are presented in **Table 2**. As shown, in this genetic background, the spreading/bunch pattern appears to correspond to a single-gene model of inheritance with no cytoplasmic effect. The allele that confers the spreading phenotype is dominant over the one that confers the bunch habit, as demonstrated by the spreading phenotype of all the F_1 hybrids, the 3:1 segregation ratio among the F_2 progeny and the 1:2:1 segregation ratio among the F_3 families. The gene was named *bunch1* and the classification of its corresponding phenotype was very easy and clear-cut, even as early as 50 days after sowing (**Figure 1**).

Bunch1 Is Associated with Several Important Agronomic Traits

In addition to growth habit, other traits with agronomic importance were examined in the segregating populations. The associations between the *bunch1* phenotype and each of these traits are presented in **Figure 2**. The bunch phenotype of *bunch1* was significantly associated with a lower dead-end ratio. The *bunch1* phenotype was also found to have a small, but significant [$\text{Prob}(t) = 0.0022$] effect on early maturation. On the other hand, the *BUNCH1* phenotype (spreading) was significantly associated with higher total pod weight and a greater number of pods per plant.

Identifying SNP Markers that Are Linked to *Bunch1*

In order to map *bunch1* on the peanut genome, a bulk segregant analysis was performed. For that analysis, 52 completely bunch and 47 completely spreading F_3 families were bulked. RNA was extracted from each bulk and converted into two libraries suitable for Illumina sequencing. After a cleaning

TABLE 2 | Segregation pattern of the spreading/bunch trait in several generations derived from closely related peanut varieties.

Generation	Hanoch (spreading) \times Harari (bunch)	Harari (bunch) \times Hanoch (spreading)
F_1	All spreading (5 plants)	All spreading (14 plants)
F_2	238 spreading: 76 bunch	194 spreading: 58 bunch
F_3 (families)	36 spreading: 81 segregating: 40 bunch	19 spreading: 37 segregating: 23 bunch

In the F_3 generation, each family contained 16 plants. Chi-square values of the 3:1 (F_2) and 1:2:1 (F_3 spreading:segregating:bunch, respectively) tests were all non-significant, indicating that data fit the expected segregation.

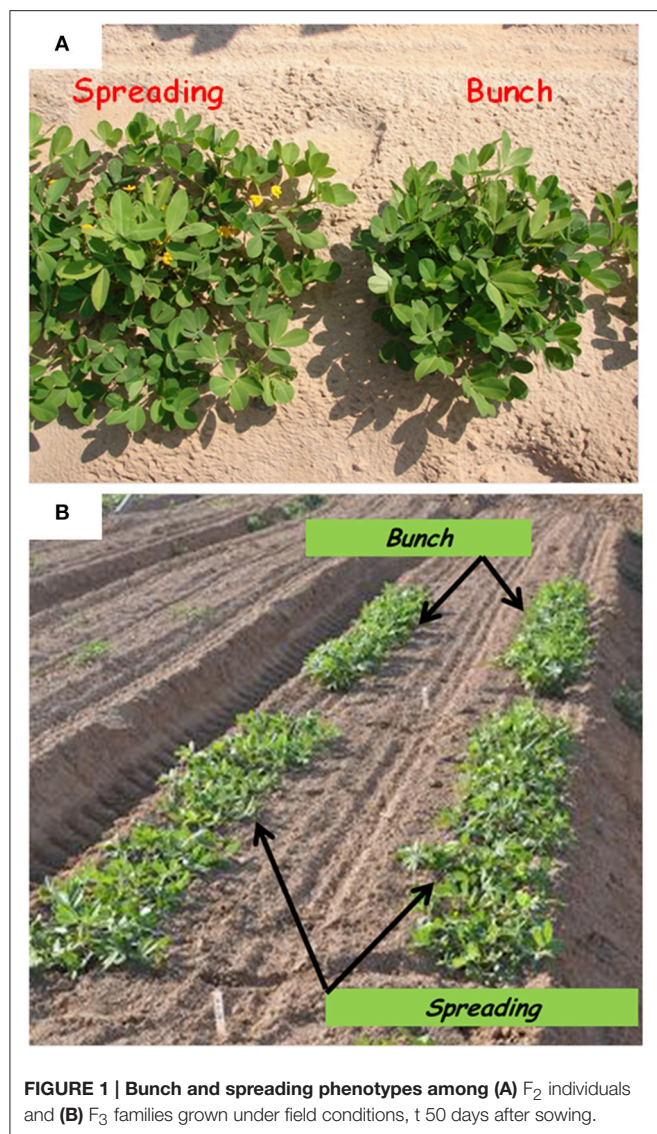


FIGURE 1 | Bunch and spreading phenotypes among (A) F_2 individuals and (B) F_3 families grown under field conditions, t 50 days after sowing.

procedure, 72 million reads per library (on average) were aligned to a 4X transcript assembly (peanutbase.org) that contains 120,364 peanut transcripts from both the A and B genomes [60,814 transcripts represent the A genome (*Arachis duranensis*) while 59,551 transcripts represent the B genome (*Arachis ipaensis*)]. With about 98% of reads mapped to the reference assembly, the expression levels of 117,957 peanut genes were measured.

Pipelines for the SNP discovery and the analysis of bulk frequency ratio were constructed according to the general scheme that was previously suggested for polyploid wheat (Trick et al., 2012). After initial filtering, ~13,000 SNPs were found to be polymorphic by the two bulks. Subsequently, the bulk frequency ratio was determined for each SNP by calculating the frequency of each nucleotide of the SNP in each bulk and then dividing one bulk by another. If the SNP is a result of false-positive call of a homoeologous SNP, or if it is not linked to the trait, then both of the SNP nucleotides will be equally

represented in the two bulks. However, if the SNP is linked to the gene, the frequencies of the SNP nucleotides in one bulk will be with significantly higher frequencies than the other bulk. In that manner, ~1,200 SNPs were found to have bulk frequency ratios of > 3 , while 34 had bulk frequency ratios > 10 (Figure 3A).

The genomic location of the 34 SNPs with bulk frequency ratios > 10 within the peanut genome was recorded (Figure 3B, Supplemental Table 1). One region at the end of linkage group 5B was found to be over-represented in this SNP group; six of the 34 were located between 5B:135,963,343 and 5B:147,304,662, including the SNP with the highest bulk frequency ratio [M875 (EZ721696.1); bulk frequency ratio = 23]. Two of these 5B linkage group and another few hypothetical SNPs with high BFR ratio from different linkage groups were further analyzed for SNP classification with Sanger sequencing. The purpose of this step was to roughly validate the location of *bunch1*. Therefore, samples from the parental lines and 6 F_2 individual plants (from which 3 spreading and 3 bunch F_3 families were derived) were selected (Supplemental Table 1). In this initial analysis, the SNPs from B5 linkage group were found almost perfectly segregating with trait, while the SNPs from the other genomic locations found to be either homoeologous SNPs (and not varietal) or didn't segregate with the trait (Supplemental Table 1), indicating for relatively high false positive ratio for the BSA by GBS technique in this system.

SNP marker M875 and other two SNPs from the same genomic location that had high bulk frequency ratios [M35 (EZ721381.1), M255 (EZ748922.1)] were further analyzed (Figure 4A). Samples from 20 homozygous F_3 families were checked (10 spreading and 10 bunch; Figure 4B). In this initial analysis, M875 was found to be completely linked to *bunch1*, while the two others, M35 and M255, were also linked to *bunch1*, but not completely. In the next phase, 182 F_4 individuals that originated from heterozygous segregating F_3 families were genotyped and phenotyped using markers M255 and M875. *Bunch1* was found to be located in a ~1.1 Mbp segment between markers M875 (B5:145,553,897; 1.9 cM) and M255 (B5:146,649,943; 2.25 cM).

Further Validation of the Bulk Segregant Analysis Using a Peanut SNP Array

Final confirmation of the fine-mapping of the *Bunch1* gene was obtained using a new Affymetrix Axiom SNP array (Pandey et al., 2017). Since the chip technology is not efficient enough to distinguish between the heterozygous and homozygous states in the polyploid, a RIL population, which was advanced from the same cv. Hanoch \times cv. Harari hybridization (F6:8), was used. Ninety-four RILs and the two parental lines were used for the analysis. Genomic DNA was extracted and applied to the 58,233 SNP clusters of the chip. Out of all of these SNPs, which were designed based on a wide spectrum of diploid and tetraploid peanut species, only 615 passed through the filtering pipeline, including significant differences between the parental lines and 1:1 segregation among the 94 RILs. The genetic analysis of these SNPs and the phenotype of *Bunch1* gene are presented

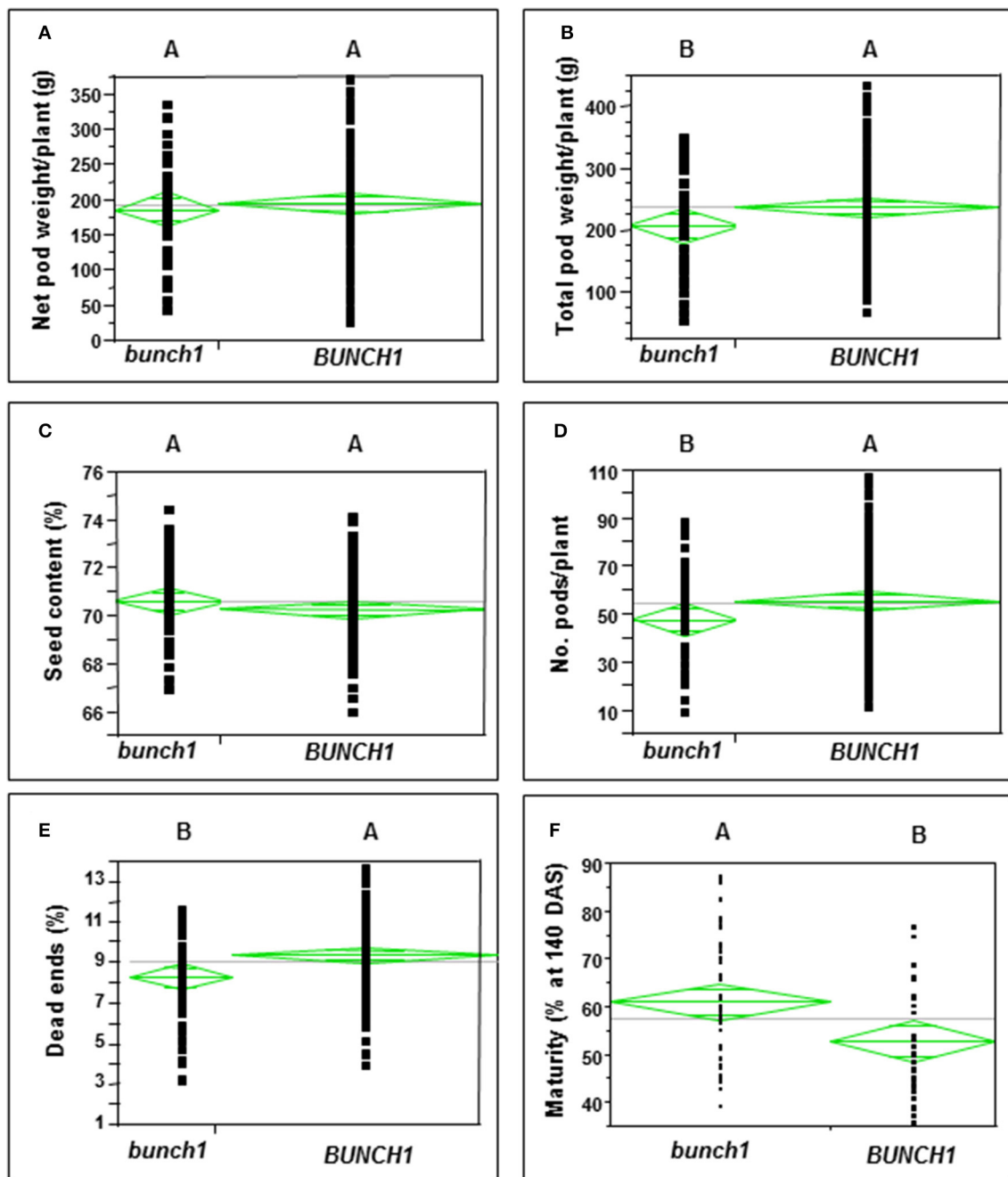


FIGURE 2 | The difference between the bunch (*bunch1*) and the spreading (*BUNCH1*) phenotypes of the *bunch1* gene in terms of several agronomic traits. Phenotypes that are not labeled with the same letter are significantly different from one another ($p < 0.05$). (A) Net pod weight, (B) total pod weight, (C) seed content, (D) number of pods per plant, (E) dead-end ratio, and (F) percentage of mature pods at 10 days after sowing (DAS).

in Figure 5A. Ten SNP markers from the array significantly ($p < 0.01$) co-segregated with the phenotype of *Bunch1* (Figure 5A). The best-linked SNP marker (AXX147251194) had only 1 recombinant RIL out of the 94 checked RILs ($p = e^{-50}$; $R^2 = 0.92$). All significant SNP markers were located in one region at linkage 5B, indicating once again that a single locus is controlling the branching habit trait in this background. These

SNPs were located in very close proximity to the three SNPs that were derived from the bulk segregant analysis (Figure 5B). Interestingly, none of the SNPs from the bulk segregant analysis were detected in the chip array and vice versa, indicating that more SNPs could possibly be found by bulk segregant analysis and used in future SNP-array designs for cultivated peanut.

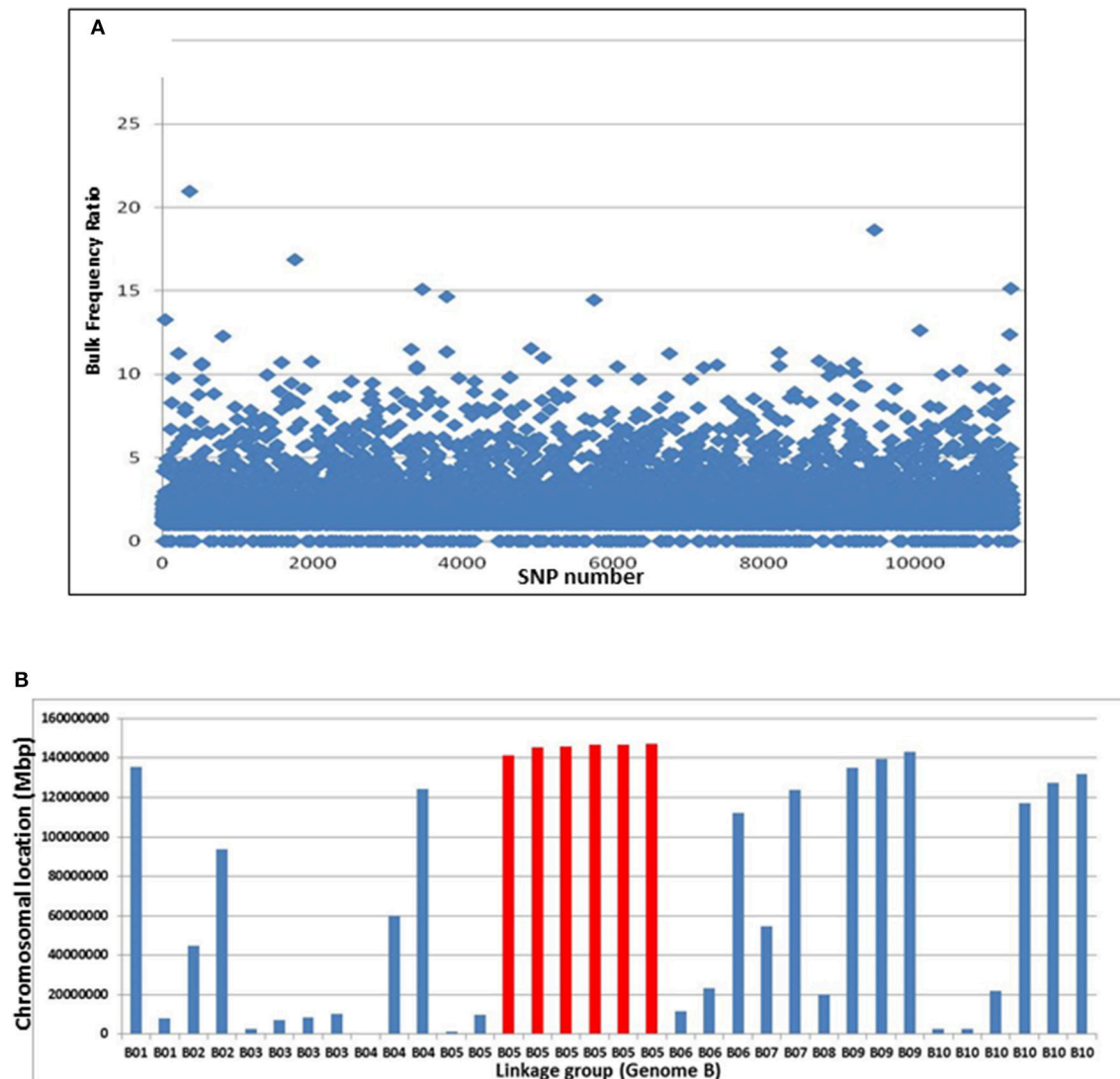


FIGURE 3 | Bulk frequency ratio (BFR) analysis to identify SNPs and genome locations that are linked to *Bunch1*. (A) The distribution of BFR in ~13,000 varietal-specific SNPs. **(B)** The genomic locations of the 34 SNPs with the highest BFR (>10), indicating one over-represented region at the end of linkage group 5B (red).

DISCUSSION

The genetic/molecular mechanism that controls growth angle in plants has been the subject of several studies, mainly involving monocotyledons, particularly rice. Several abnormal tiller-angle mutants and their corresponding genes have been reported in rice, such as *LA1* (Li et al., 2007) and *PIN2* (Chen et al., 2012). Two additional genes with opposite effects on tiller angle, *Tiller Angle Control 1* (*TAC1*) and *Prostrate Growth 1* (*PROG1*), have also been identified in rice (Yu et al., 2007; Tan et al., 2008). These genes have played critical roles in the domestication of rice. There are several reports regarding the molecular biology of the growth angle of lateral shoots of dicot species. Roychoudhry et al. (2013) described a model in which the set point angle of lateral

branches of higher plants is controlled by an auxin-dependent antigravitropic mechanism. The molecular basis for the spatial pattern of tree branches was also studied in peach, resulting in the discovery of a new ortholog of the *TAC1* gene, which controls the “pillar” tree phenotype (Dardick et al., 2013).

We explored the branching habit in the leguminous crop *Arachis hypogaea* and fine-mapped a major gene that controls this trait. The *bunch1* gene was mapped to a relatively small genomic region that includes ~70 ORFs for gene models. Interestingly, BlastX analysis showed that none of the above-mentioned genes that control the growth angle in either monocots or dicots were present in the peanut genome or mapped in approximation to *bunch1*. The genetic controller of *bunch1* may therefore be novel. Several candidate genes involved in plant hormone metabolism

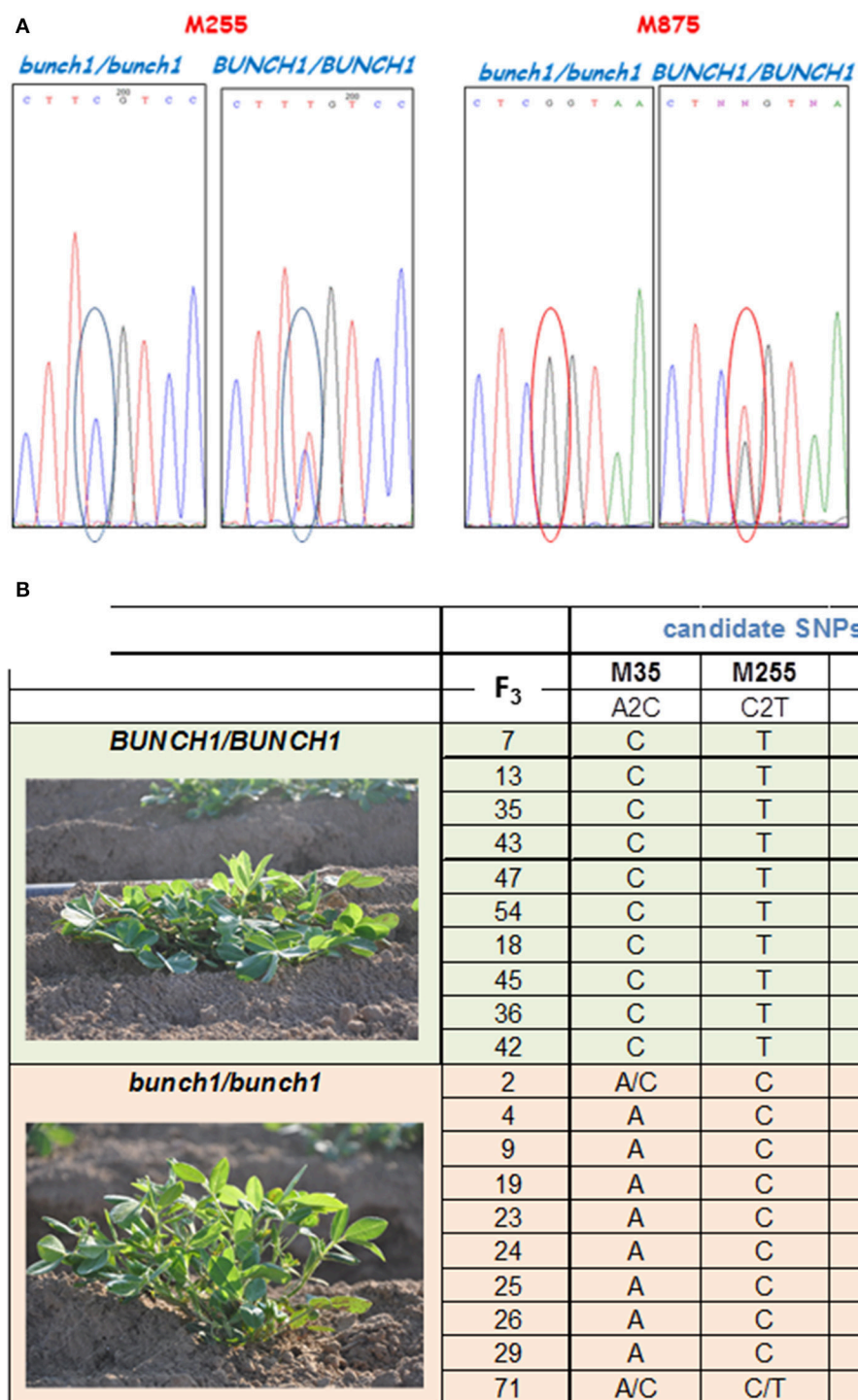


FIGURE 4 | SNP validation and linkage analysis with *Bunch1*. (A) An example of Sanger sequencing validation of two SNP markers. (B) Analysis of linkage between the top three SNP markers and the phenotype of *Bunch1*. Samples from 10 completely spreading (*BUNCH1/BUNCH1*) and 10 completely bunch (*Bunch1/Bunch1*) F₃ families were analyzed.

and light reception are located within that region and have been identified as possibly controlling *bunch1*. One of these may be a *FAR1-Related* sequence (B05:146200756..146203528) that

encodes a family of proteins that are essential for phytochrome A-controlled far-red responses in *Arabidopsis thaliana*; (Lin and Wang, 2004). Another putative candidate gene

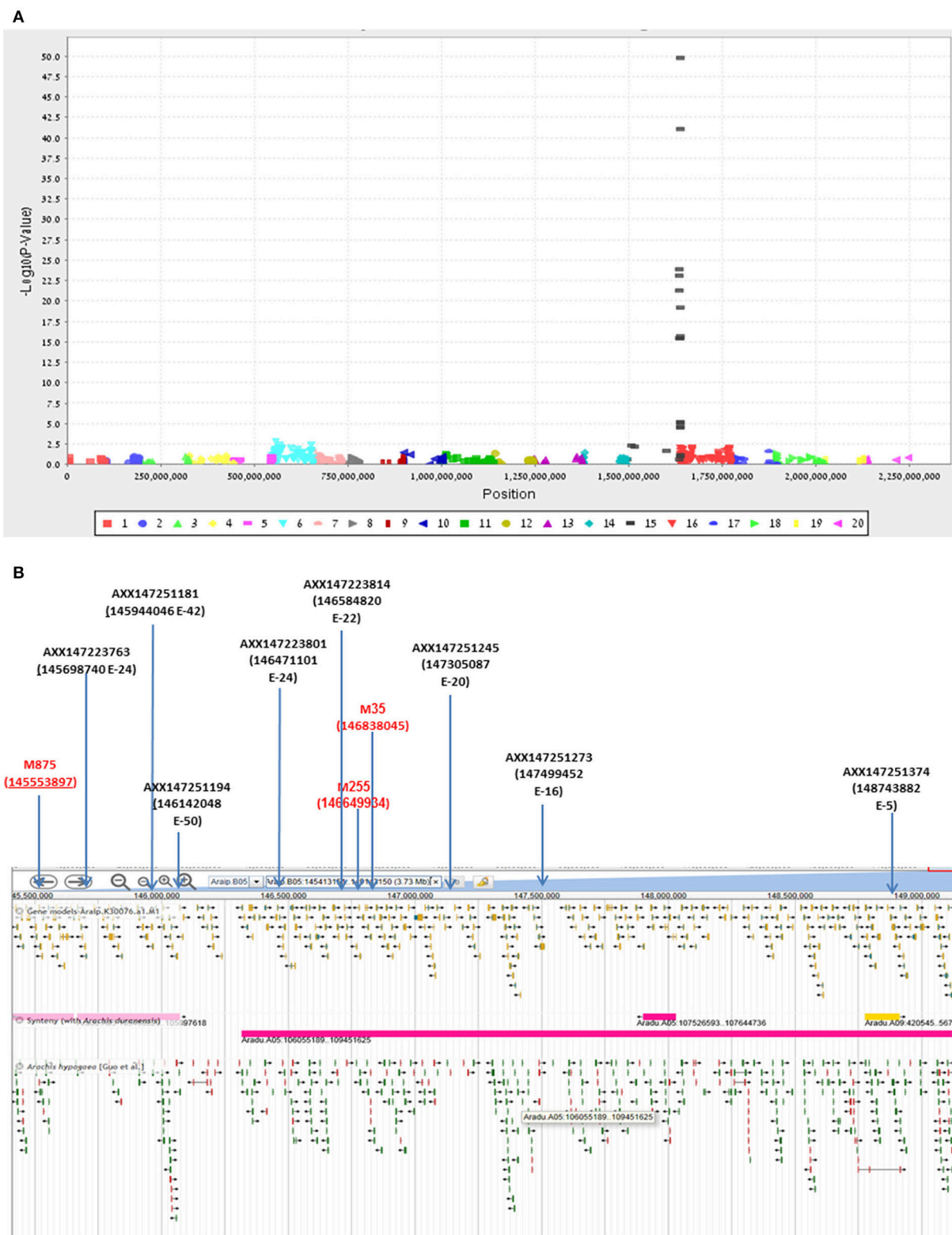


FIGURE 5 | Validation and further fine-mapping of *Bunch1* using an Affymetrix Axiom SNP array. (A) Manhattan plot for the analysis of linkage between 615 Affymetrix Axiom SNPs and the *Bunch1* phenotype. 1–10 = genome A; 11–20 = genome B (e.g., 15 = linkage group 5B). **(B)** Integrative map for the bulk segregant analysis and SNP-array analyses of the peanut linkage group B5 (from PeanutBase.org). Markers derived from the bulk segregant analysis are indicated in red. Yellow—gene models. Green—ESTs of genes. Pink—synteny of this region with *A. duranensis* (Genome A).

is the 1-aminocyclopropane-1-carboxylate oxidase-like protein (ACC-oxidase; B05:146236653..146238358), which catalyzes the last step in ethylene biosynthesis. Ethylene biosynthesis may play an important role in determining peanut growth angle. Applying a relatively small amount of EPCA (an ethylene-releasing compound) caused the horizontal branches of runner-type plants to become erect (Ziv et al., 1976). Yet, these and other candidate genes must, of course, be further examined in light- and plant hormone-targeted studies, as well as subjected to verification by positional cloning and transformation.

The *Bunch1* gene had strong associations with several traits in the segregating F₂ and F₃ populations. Plants with the bunch phenotype had, on average, earlier maturity values and fewer dead-end pods. Plants with the spreading phenotype had on average more pods per plant, but many of those pods were actually undeveloped. In the bunch type, especially when a wide planting spacing is used (as in our experiments), many flowers are too far from the ground and cannot reach to the soil. For that manner, only the pods that are close to the root will develop. However, those pods that reach the soil develop in a more synchronized manner among the bunch types than among the spreading types. This promotes uniform maturation, better pod filling and, eventually, fewer dead-end pods. Indeed, many of the dead-end pods in the spreading types (like cv. Hanoch) are from distal parts of the branches, where pods develop late in the season and do not fully mature. In a subset of the RIL population, the bunch phenotype of the branching gene is significantly correlated with greater resistance to white mold (caused by *S. sclerotiorum*; data not shown; submitted for publication). Therefore, it is suggested that the phenotype of the *Bunch1* gene has an important agricultural role in Virginia-type peanuts.

As we explored the segregation patterns of several other crosses between Virginia-type related cultivars with different branching habits, we noticed that this model of the *bunch1* gene is relatively common within Israeli peanut breeding germplasm. Crosses between cv. Hanoch (spreading) and cv. Hillah/Shulamit (bunch) resulted in a 3:1 spreading:bunch ratio (data not shown), tracing the origin of this trait back to the early 1970s. Crosses between cv. Harari (bunch) and a runner-type peanut line GK-7-OL (spreading) also resulted in a 3:1 ratio (data not shown), indicating that the single-gene model for this trait is not confined to the Virginia-type germplasm. Yet, in crosses with Valencia-type peanut germplasm (plants with an erect branching habit), this system of single-gene inheritance was not found and the branching habit was therefore relatively hard to classify. We conclude that the *bunch1* phenotype is confined to the *A. hypogaea* ssp. *hypogaea* gene pool. However, other allelic variations of this gene may exist in the *A. hypogaea* ssp. *fastigiata* germplasm, since some of the bunch-habit Virginia-type lines in

Israel (e.g., cvs. Shulamit, Hillah, etc.) have *A. fastigiata* origins. Moreover, the branching habit trait of peanut was analyzed in a previous study by Fonceka et al. (2012a), which involved mapping traits in a cross between amphidiploid *A. ipaensis* /*A. duranensis* and a Spanish-type cultivar (*A. hypogaea* ssp. *fastigiata* var *vulgaris*) with an erect growth habit. There was one significant QTL for the branching habit at the same location as *Bunch1* (at the end of linkage group B5), explaining 16.2 of the total variation of the trait in the population. It is very likely that the reported QTL for branching habit and the locus of *Bunch1* are the same. This demonstrates once again that the origin of the *Bunch1* phenotype may be beyond the *A. hypogaea* ssp. *hypogaea* genetic background.

We have demonstrated the relatively straight-forward and easy utilization of bulk segregant analysis for the fine-mapping of a monogenic trait in the low-polymorphic and heterozygous germplasm of cultivated peanut. Also, although the bunch/spreading trait is very easy to classify, there may be some new uses for these new markers in peanut breeding, particularly in the validation of successful F₁ hybrids (when the female of the cross is the dominant spreading type) and the selection of homozygous spreading families in early breeding generations. The fine-mapping of this trait also provides a baseline for the cloning of *Bunch1*, presumably one of the first map-based positional cloning ventures in current genetic research of peanut.

AUTHOR CONTRIBUTIONS

GK and YB were responsible for the molecular work; AF was responsible for the bioinformatic analysis; AP was responsible for RIL population analysis; IH conducted the field trials; RH managed the study and wrote the manuscript.

ACKNOWLEDGMENTS

This study was funded by the Israel Peanut Production and Marketing Board. The authors wish to thank Mr. Oren Buchshtab and other workers of Dod Moshe LTD for their assistance with the field trials.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00467/full#supplementary-material>

Supplemental Table 1 | List of the 34 hypothetical SNPs with the highest (>10) BFR. The table indicates the location of the SNP, the change of the nucleotide and the statistic score for the change. Also, the table presents further Sanger sequencing analysis of six SNPs with high BFR ratio. For this, six F₂ individual plants were sampled, from which three spreading (green) and three bunch (yellow) F₃ families were derived.

REFERENCES

- Ashri, A. (1964). Intergenic and genic-cytoplasmic interactions affect growth habit in peanuts. *Genetics* 50, 363–372.
- Ashri, A. (1968). Genic-cytoplasmic interactions affect growth habit in peanut, *A. hypogaea* II. A revised model. *Genetics* 60, 807–810.
- Ashri, A. (1975). *Natural Cytoplasmic Divergence and Induction of Plasmion Mutations in Peanuts (Arachis hypogaea)*. Final Res Report, USDA, ARS, PL480, Project No AID-CR-77 Grant No FG-IS 218, 1–32.
- Ashri, A., and Levi, A. (1975). Natural and induced variations affecting growth habit in peanuts. *J. Am. Peanut. Res. Educ. Assoc.* 7:64.

- Bertioli, D. J., Cannon, S. B., Froenicke, L., Huang, G. D., Farmer, A. D., Cannon, E. K., et al. (2016). The genome sequences of *Arachis duranensis* and *Arachis ipaensis*, the diploid ancestors of cultivated peanut. *Nat. Genet.* 48, 438–444. doi: 10.1038/ng.3517
- Brand, Y., and Hovav, R. (2010). Identification of suitable internal control genes for quantitative Real-Time PCR expression analyses in peanut (*Arachis hypogaea*). *Peanut Sci.* 37, 12–19. doi: 10.3146/PS09-014.1
- Butzler, T. M., Bailey, J., and Beute, M. K. (1998). Integrated management of Sclerotinia blight in peanut: utilizing canopy morphology, mechanical pruning, and fungicide timing. *Plant Dis.* 82, 1312–1318. doi: 10.1094/PDIS.1998.82.12.1312
- Chen, Y. N., Fan, X., Song, W., Zhang, Y., and Xu, G. (2012). Over-expression of OsPIN2 leads to increased tiller numbers, angle and shorter plant height through suppression of OsLAZY1. *Plant Biotechnol. J.* 10, 139–149. doi: 10.1111/j.1467-7652.2011.00637.x
- Clevenger, J., Chu, Y., Chavarro, C., Agarwal, G., Bertioli, D. J., Leal-Bertioli, S. C., et al. (2016). Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Mol. Plant* 10, 309–322. doi: 10.1016/j.molp.2016.11.015
- Coffelt, T. A. (1974). Inheritance of growth habit in an intra-specific cross population of peanuts. *J. Hered.* 65, 160–162. doi: 10.1093/oxfordjournals.jhered.a108489
- Dardick, C., Callahan, A., Horn, R., Ruiz, K. B., Zhebentyayeva, T., Hollender, C., et al. (2013). PpTAC1 promotes the horizontal growth of branches in peach trees and is a member of a functionally conserved gene family found in diverse plants species. *Plant J.* 75, 618–630. doi: 10.1111/tpj.12234
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K., V., Maguire, J., R., Hartl, C., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Faostat (1998). Available online at: <http://faostat.fao.org/default.aspx>, last update 2016.
- Fonceka, D., Tossim, H. A., Rvallan, R., Vignes, H., Faye, I., Ndoeye, O., et al. (2012a). Fostered and left behind alleles in peanut: interspecific QTL mapping reveals footprints of domestication and useful natural variation for breeding. *BMC Plant Biol.* 12:26. doi: 10.1186/1471-2229-12-26
- Fonceka, D., Tossim, H. A., Rvallan, R., Vignes, H., Lacut, E., de Bellis, F., et al. (2012b). Construction of chromosome segment substitution lines in peanut (*Arachis hypogaea* L.) using a wild synthetic and QTL mapping for plant morphology. *PLoS ONE* 7:e48642. doi: 10.1371/journal.pone.0048642
- Gupta, K., Buchshtab, O., and Hovav, R. (2014). The effects of irrigation level and genotype on pod-filling related traits in peanut (*Arachis hypogaea*). *J. Agric. Sci.* 7, 169–181. doi: 10.5539/jas.v7n1p169
- Gupta, K., Hedvat, I., Faigenboim-Doron, A., Clevenger, J., Ozias-Akins, P., and Hovav, H. (2016). Transcriptome profiling of peanut developing seed with a focus on duplicate oil related pathways. *Front. Plant Sci.* 7:1491. doi: 10.3389/fpls.2016.01491
- Hull, F. H. (1933). The classification of groundnut varieties with a preliminary note on inheritance of some characters. *Trop. Agric. West Indies* 70, 318–327.
- Kochert, G., Halward, T., Branch, W. D., and Simpson, C. E. (1991). Rflp variability in peanut (*Arachis-Hypogaea* L.) cultivars and wild-species. *Theor. Appl. Genet.* 81, 565–570. doi: 10.1007/BF00226719
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 354–357. doi: 10.1038/nmeth.1923
- Li, P., Wang, Y., Qian, Q., Fu, Z., Wang, M., Zeng, D., et al. (2007). LAZY1 controls rice shoot gravitropism through regulating polar auxin transport. *Cell Res.* 17, 402–410. doi: 10.1038/cr.2007.38
- Lin, R., and Wang, H. (2004). Arabidopsis FHY3/FAR1 gene family and distinct roles of its members in light control of arabidopsis development. *Plant Physiol.* 136, 4010–4022. doi: 10.1104/pp.104.052191
- Magwene, P. M., Willis, J. H., and Kelly, J. K. (2011). The statistics of bulk segregant analysis using next generation sequencing. *PLoS Comput. Biol.* 7:e2255. doi: 10.1371/journal.pcbi.1002255
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Moretzsohn, M. C., Gouvea, E. G., Inglis, P. W., Leal-Bertioli, S. C., Valls, J. F., and Bertioli, D. J. (2013). A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Ann. Bot. Lond.* 111, 113–126. doi: 10.1093/aob/mcs237
- Ozias-Akins, P. (2013). The orphan legume genome whosetime has come: symposium highlights from the American Peanut Research & Education Society annual meeting. *Peanut Sci.* 40, 66–69. doi: 10.3146/PS13-14.1
- Pandey, M., Agarwal, G., Kale, S., Clevenger, J., Nayak, S., Sriswathi, M., et al. (2017). Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7:40577. doi: 10.1038/srep40577
- Patel, J. S., John, C. M., and Seshadri, C. R. (1936). The inheritance of characters in groundnut, *A. hypogaea*. *Proc. Ind. Acad. Sci.* 3, 214–233.
- Pittman, R. N. (1995). *United States peanut Descriptors*, ARS 132. Tifton, GA: U.S. Department of Agriculture, Agricultural Research Service.
- Roychoudhry, S., Del Bianco, M., Kieffer, M., and Kepinski, S. (2013). Auxin controls gravitropic setpoint angle in higher plant lateral branches. *Curr. Biol.* 23, 1497–1504. doi: 10.1016/j.cub.2013.06.034
- Selvaraj, M. G., Narayana, M., Schubert, A. M., Ayers, J. L., Baring, M. R., and Burrow, M. D. (2009). Identification of QTLs for pod and kernel traits in cultivated peanut by bulked segregant analysis. *Electron. J. Biotechnol.* 12. doi: 10.2225/vol12-issue2-fulltext-13
- Tan, L. B., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., et al. (2008). Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* 40, 1360–1364. doi: 10.1038/ng.197
- Trick, M., Adamski, N. M., Mugford, S. G., Jiang, C. C., Febrer, M., and Uauy, C. (2012). Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* 12:14. doi: 10.1186/1471-2229-12-14
- Yu, B., Lin, Z., Li, H., Li, X., Li, J., Wang, Y., et al. (2007). TAC1, a major quantitative trait locus controlling tiller angle in rice. *Plant J.* 52, 891–898. doi: 10.1111/j.1365-313X.2007.03284.x
- Zhou, X., Xia, Y., Ren, X., Chen, Y., Huang, L., Huang, S., et al. (2014). Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). *BMC Genom.* 15:351. doi: 10.1186/1471-2164-15-351
- Ziv, M., Koller, D., and Halevy, A. H. (1976). Ethylene and the geotropic response of lateral branches in peanuts (*Arachis hypogaea* L.). *Plant Cell Physiol.* 17, 333–339.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Kayam, Brand, Faigenboim-Doron, Patil, Hedvat and Hovav. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Phenotyping: Using Machine Learning for Improved Pairwise Genotype Classification Based on Root Traits

Jiangsan Zhao¹, Gernot Bodner² and Boris Rewald^{1*}

¹ Department of Forest and Soil Sciences, University of Natural Resources and Life Sciences, Vienna, Austria, ² Division of Agronomy, Department of Crop Sciences, University of Natural Resources and Life Sciences, Vienna, Austria

OPEN ACCESS

Edited by:

Maria Carlota Vaz Patto,
ITQB-Universidade Nova de Lisboa,
Portugal

Reviewed by:

Anjali Iyer-Pascuzzi,
Purdue University, USA
Joao Mendes Mendes-Moreira,
University of Porto, Portugal

*Correspondence:

Boris Rewald
rewald@rootecology.de

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 25 July 2016

Accepted: 25 November 2016

Published: 06 December 2016

Citation:

Zhao J, Bodner G and Rewald B
(2016) Phenotyping: Using Machine
Learning for Improved Pairwise
Genotype Classification Based on
Root Traits. *Front. Plant Sci.* 7:1864.
doi: 10.3389/fpls.2016.01864

Phenotyping local crop cultivars is becoming more and more important, as they are an important genetic source for breeding – especially in regard to inherent root system architectures. Machine learning algorithms are promising tools to assist in the analysis of complex data sets; novel approaches are needed to apply them on root phenotyping data of mature plants. A greenhouse experiment was conducted in large, sand-filled columns to differentiate 16 European *Pisum sativum* cultivars based on 36 manually derived root traits. Through combining random forest and support vector machine models, machine learning algorithms were successfully used for unbiased identification of most distinguishing root traits and subsequent pairwise cultivar differentiation. Up to 86% of pea cultivar pairs could be distinguished based on top five important root traits (Timp5) – Timp5 differed widely between cultivar pairs. Selecting top important root traits (Timp) provided a significant improved classification compared to using all available traits or randomly selected trait sets. The most frequent Timp of mature pea cultivars was total surface area of lateral roots originating from tap root segments at 0–5 cm depth. The high classification rate implies that culturing did not lead to a major loss of variability in root system architecture in the studied pea cultivars. Our results illustrate the potential of machine learning approaches for unbiased (root) trait selection and cultivar classification based on rather small, complex phenotypic data sets derived from pot experiments. Powerful statistical approaches are essential to make use of the increasing amount of (root) phenotyping information, integrating the complex trait sets describing crop cultivars.

Keywords: breeding, cultivar classification, pea (*Pisum sativum* L.), random forest (RF), root phenotyping, root trait selection, support vector machine (SVM)

INTRODUCTION

A policy report of the European Union noted recently that protein crops, e.g., bean, lentil, lupine, pea, and soya, are currently grown on 1.8% of arable land in the EU only, compared with 4.7% in 1961, and about 8% in Australia and Canada, 14.5% in North America, and 25.5% in South America (Jiang et al., 2004; EU, 2013; FAOSTAT, 2014; Cernay et al., 2015). This is despite grain legumes representing a significant source of protein for food (Vaz Patto et al., 2014; Multari et al., 2015).

and feed (Jezierny et al., 2010; Koivunen et al., 2014), and legume cultivation reducing the need for N fertilizer even for subsequent crops in the rotation (Preissel et al., 2015). Recent studies have identified a comparative lack of breeding investment in Europe to improve grain legume adaptation to local agro-climatic conditions and management techniques (Annicchiarico and Iannucci, 2008; Lizarazo et al., 2015). While distinct leguminous crops are used locally for food and feed, and local cultivars are kept in numerous collections in gene banks, research institutions, and also in farms/home gardens, this genetic pool cannot be used at its full potential for large-scale agriculture and breeding programs until important traits have been determined.

Plant phenotyping intends measuring complex traits related to growth, yield, and adaptation to stress at different macroscopically scales of plant organization (Fiorani and Schurr, 2013). Examples for measured parameters are leaf vein structure (Sack and Scoffoni, 2013; Caringella et al., 2015), photosynthetic efficiency (Gilbert et al., 2011; Gorbe and Calatayud, 2012; Grady et al., 2013), root morphology (Iyer-Pascuzzi et al., 2010; Bucksch et al., 2014), biomass (Stachowicz et al., 2013; Poorter et al., 2015), and yield quantity and quality (Iqbal and Lewandowski, 2014).

Especially the modification of root system architecture (RSA) could contribute to improvements of desirable agronomic traits such as yield, drought tolerance, and resistance to nutrient deficiencies; thus, RSA was described key to a second green revolution improving resource use efficiency of crops (Lynch, 2007). For example, an architectural trait enhancing topsoil foraging is a higher number of basal roots – contributing significantly to phosphorous acquisition (Lynch, 2011), while deep rooting can, e.g., sustain water acquisition during drought periods or improve uptake of percolating nitrate (Pinheiro et al., 2005). Recently, an increased focus was laid on improving high-throughput, image-based root phenotyping approaches (Berger et al., 2010; Hartmann et al., 2011; Mairhofer et al., 2012; Fiorani and Schurr, 2013; Li et al., 2014). See Kuijken et al. (2015) for a recent review on the latest developments in root phenotyping and an overview on environmental and genetic factors influencing root phenotypes.

Advanced machine learning (ML) approaches encompass promising statistical tools for variable selection and group classification. While the use of ML approaches in non-genetic/biochemical plant sciences is still scarce, ML was introduced to promote RSA classification in the recent past (Zhong et al., 2009; Iyer-Pascuzzi et al., 2010). Iyer-Pascuzzi et al. (2010) use of ML methods mainly benefited from using a non-invasive imaging system which enabled them to acquire 16 traits from a high number of pictures [~200 (pseudo-)replicates per genotype]. However, the limited plant age and highly artificial growth conditions are major disadvantages of many non-invasive and high-throughput root phenotyping methods (Bucksch et al., 2014): RSA differs with ontogeny (Hargreaves et al., 2009; Wojciechowski et al., 2009; Hund et al., 2011) and is highly plastic to edaphic conditions (Tracy et al., 2012; Rich and Watt, 2013). Thus, analyses on mature plants *in situ* or under more realistic growth conditions, which mostly rely on manual, destructive methods [e.g., 'shovelomics'; Trachsel et al. (2011)] continue to be

essential although the number of replicates and measurements is often limited but the complexity of variables (i.e., traits) remains high. Similar to the situation in cell biology (Sommer and Gerlich, 2013), available ML approaches in plant sciences have been optimized for large-scale screenings, probably partially due to the difficulty in applying ML algorithms with unbiased variable selection on low number of replicates (Bucksch et al., 2014).

Among supervised ML algorithms, random forest (RF) is a non-parametric method with high accuracy and robustness to noise (Breiman, 2001). RF has been applied in several biological fields, like gene (Díaz-Uriarte and De Andres, 2006) and protein sequence (Pan and Shen, 2009) selection, and disease prediction (Yang et al., 2014). However, most previous studies mainly focused on improving classification accuracy with variable selection rather than variable interpretation (Liu et al., 2014; da Costa et al., 2015; Gowin et al., 2015) because the variable importance measure is biased in the standard RF algorithm – overestimating the importance of correlated predictor variables (Strobl and Zeileis, 2008). However, unbiased variable selection is essential to stable classification and meaningful interpretation of plant traits and other data and can be achieved by using an improved RF algorithm – based on a conditional permutation scheme as a computational means to determine variable importance (Strobl and Zeileis, 2008; Strobl et al., 2008). Support vector machines (SVMs) are another set of supervised ML methods which can be trained to classify individuals in high-dimension space (Cortes and Vapnik, 1995). SVMs have been widely used in neuro-image classification (Gaonkar and Davatzikos, 2013) and face detection (Shan, 2012). SVMs can be differentiated based on kernel functions (Okkan and Serbes, 2012): linear kernel functions (linear SVMs) were previously used for variable selection of root systems (Iyer-Pascuzzi et al., 2010). However, variable selection by ranking absolute values of weights are biased, as the absolute weight values of irrelevant variables can be as high as of important ones (Statnikov et al., 2006; Gaonkar and Davatzikos, 2013). SVMs based on Gaussian radial basis function (rbf) kernel often provide a better performance on 'noisy' data sets not separable linearly – resulting in the more widely use of rbf SVMs in classification (Hsu et al., 2010).

Powerful statistical approaches are essential to make use of the increasing amount of (root) phenotyping information, integrating the complex trait sets (describing RSA). Combining RF with rbf SVMs for variable selection and group classification, respectively, might overcome problems in applying ML approaches on data sets characterized by a rather low signal-to-noise ratio such as manually derived phenotyping data (Liu et al., 2004). For example, Löw et al. (2012) found a significantly higher classification accuracy of pre-crops in two out of four agricultural regions using satellite images when applying a combination of RF and SVMs compared to using either RF or SVMs for both trait selection and classification. Thus, aims of this study were to determine if (i) RF can be reproducibly used for selecting important root traits (i.e., root traits distinguishing mature pea cultivars), and (ii) how root trait selection influences cultivar classification by rbf SVMs. We hypothesize that rbf SVMs classification is superior to traditional univariate tests if important root traits are identified by RF. *Pisum sativum* L. was

selected as test species because it is one of the most frequently cultivated grain legumes worldwide (Alves-Carvalho et al., 2015), with especially European genetic resources still insufficiently characterized.

MATERIALS AND METHODS

Plant Material and Experimental Set-Up

Sixteen randomly selected cultivars of pea (*P. sativum* L.) were used for root phenotyping (Table 1), originating either from Southern (Portugal and Spain) or Northern Europe (Estonia, Latvia, Norway, and Sweden). Seeds were provided by partners within the EU FP7 project 'Eurolegume' and by the Nordic gene bank.

Experiments were conducted in a large plastic foil greenhouse from June 13th, 2014 to October 7th, 2014 located in Tulln, Austria (48.33°N, 16.05°E). Aeration openings of the greenhouse were fitted with mesh to prevent insect infestations. Solar radiation, air temperature, and relative humidity were hourly recorded 2 m above ground; mean air temperature during the measurement period was 20.7°C, relative humidity ranged between 18.6–87.7% with a mean of 59.9%. Mean daily sum of solar radiation was 13.28 MJ m⁻² day⁻¹, with a maximum of 27.41 MJ m⁻² day⁻¹ at July 1st; day length (solar radiation ≥ 120 W m⁻²) varied between 10 and 16 h.

TABLE 1 | Sixteen pea (*Pisum sativum* L.) cultivars used locally for food in different European countries and institutions donating the seeds for the experiment.

Abbreviation	Cultivar	Country of origin	Donor institution
ps1.Estonia	Eesti hall	Estonia	ECRI
ps2.Estonia	Eesti kollane soogihernes	Estonia	ECRI
ps3.Estonia	Jõgeva roheline	Estonia	ECRI
ps4.Estonia	Seko	Estonia	ECRI
ps5.Latvia	Alma	Latvia	SPPBI
ps6.Latvia	Bruno	Latvia	SPPBI
ps7.Latvia	k-4833 Stendes Hero	Latvia	SPPBI
ps8.Latvia	Retrija	Latvia	SPPBI
ps9.Norway	NGB 10778.1	Norway	NordGen
ps10.Norway	NGB 20045.3	Norway	NordGen
ps11.Portugal	Gp 3263	Portugal	INIAV
ps12.Portugal	Gp 3491	Portugal	INIAV
ps13.Portugal	Gp 3497	Portugal	INIAV
ps14.Portugal	Grisel	Portugal	INIAV
ps15.Sweden	NGB 1025131	Sweden	NordGen/ Uppsala U
ps16.Sweden	NGB 131381	Sweden	NordGen/ Uppsala U

ECRI, Estonia Crop Research Institute; NordGen, Nordic Genetic Resource Center, Norway; INIAV, Instituto Nacional de Investigação Agrária e Veterinária, Portugal; SPPBI, Priekuli Plant Breeding Institute, Latvia; Uppsala U, Uppsala University, Sweden.

Seeds of all cultivars were germinated in a growth chamber (Fitotron; Weiss-Gallenkamp, UK) at 25 ± 1°C. Seeds were coated with a rhizobium suspension (Steinberga et al., 2008) before being planted in 0.5-L plastic bags (10 cm high) filled with washed quartz sand (0.7–1.2 mm-sized) amended with 1 g of slow release fertilizer (Osmocote Pro 3-4M; 17-11-10+2MgO+TE; ICL Specialty Fertilizers, Tel Aviv, Israel). Initial germination was conducted in darkness; after the first seed germinated, light (PAR 350 μE m⁻² s⁻¹) was turned on (16 h light/8 h dark). Germination time varied between 4 and 6 days with minor differences between cultivars (data not shown). After 10–14 days, eight similar-sized seedlings per cultivar were selected for transplanting.

In the greenhouse, eight blocks of 16 plastic tubes each (128 tubes in total) were established on wooden frames in North to South direction. The plastic tubes used as pots/growing cylinders in the experiments were 108 cm long and 20 cm in diameter (~32 L); the bottom was sealed with a cap; holes covered with a glass fiber mat allowed for free drainage. Before the tubes were filled with washed, 0.7–1.2 mm-sized quartz sand, a plastic liner was installed in each tube allowing for undisturbed removal of the substrate during harvest; the liner was perforated at the bottom 10 cm. Measurements in large plastic tubes have previously shown good agreement with maximum rooting depth and root length density as determined in the field and have been used to explore root traits in other legumes such as chickpea (Kashiwagi et al., 2006; Vadez et al., 2008). For transplanting, germination bags were placed inside the tubes and cut open at the side and bottom to prevent root disturbance. One plant per cultivar was randomly arranged in each of the eight blocks. 8.3 g of an AMF inoculum (*Glomus mosseae* BEG95, *G. intraradices*, and *G. geosporum* BEG199; supplied by Dr. Aleš Látr, Symbiom, Czech Republic) were added to each plant individual around the root systems at depths of 0–10 cm before the tube was brimmed with additional sand. An automated, pressure-compensated drip-irrigation system was used to supply all plants with ample amounts of water and a modified Long Ashton nutrient solution (Jia et al., 2004); amounts were adjusted to increasing plant size and weather conditions.

Harvest and Analysis

Plants were randomly harvested within blocks at 71–92 days after transplanting. After harvesting the shoots (data not shown), the tubes were placed horizontally and the plastic liner was pulled out on a 1.5 mm-mesh table. After the plastic liner was cut open, roots were then manually excavated as previously described by Kashiwagi et al. (2005) and others. No roots reached the bottom of the tube and few roots were discovered at the sides, indicating a rather unrestricting pot size. After the root system was uncovered, the maximum rooting depth was determined. It was further washed and rinsed in a bucket filled with clean tap water (Miguel, 2012), photographed next to a size standard, stored in a water-filled plastic bag, and transported to the lab for further analysis. Detached (i.e., shed/broken off) root segments were accurately collected from the remaining sand on the mesh table (mesh size: 2 × 2 mm), stored in paper bags and transported to the lab, oven dried (65°C, 48 h) and added to root biomass. In

the lab, the root systems were stored at 4°C until further analysis (≤ 3 weeks) took place (Hu et al., 2013).

For in-depth architectural and morphological analysis, the root systems of 5–7 plant individuals per cultivar (97 plants in total) were manually dissected into tap root and laterals. Laterals along the tap root and the tap root were separated into the five depth classes 0–5, 5–10, 10–20, 20–40, and 40–100 cm. Three lateral root samples from the depth classes 0–5, 5–10, and 10–20 cm along the tap root were scanned in water-filled trays (Epson Expression 10000XL; Epson, Nagano, Japan) at 400 dpi, grayscale. Pictures were analyzed for diameter, surface area, length, and volume with the PC program WinRhizo 2012b Pro (Régent Inst., Ville de Québec, QC, Canada). Subsequently, all root samples were dried (65°C, 48 h) and weighed to an accuracy of ± 0.1 mg (CP225D; Sartorius, Göttingen, Germany). The specific root area (SRA, $\text{cm}^2 \text{g}^{-1}$), specific root length (SRL, cm g^{-1}), tissue density (TD, g cm^{-3}), the total root surface area (totalRSA), root length (RL), and root volume were calculated. Determined root traits (Guo et al., 2008; Rewald et al., 2011) are listed in **Table 2** (Cramer et al., 2007; Alves-Carvalho et al., 2015).

Data Analysis

Thirty-six root traits in total, either directly measured or calculated, were available for analysis (**Table 2**). Non-normal distributed root traits were box-cox transformed with ‘MASS’ package, version 7.3-44, in R for Windows version 3.2.2 (R Core Team, 2015). Multiple imputation was conducted by ‘Amelia’ package, version 1.7.3. The R code used for data preparation can be found as Supplementary Method S1.

Random Forest and Support Vector Machines

Random forest (Breiman, 2001; Strobl et al., 2009) was used to rank root traits according to their importance for classification; SVMs (Vapnik and Vapnik, 1998) were used for multiclass or pairwise cultivar classification. A flow chart outlining data handling steps can be found as Supplementary Method S2.

Because the multiclass classification resulted in very low accuracy (see Supplementary Figure S1), only the pairwise classification was pursued further. In order to use the ‘cforest’ function in the R ‘Party’ package (version 1.0–23) for root traits importance measure, individuals of each pair of cultivars were oversampled four-times (to gain the number of data points required by the software algorithm) for pairwise classification. As RF was used for traits important measure only, the whole oversampled data set was used. Afterward, the number of root traits randomly chosen at each split (building each tree), *mtry*, was tuned. Even though it has been suggested that $mtry = \sqrt{n}$, n is the amount of root traits, always generates acceptable classification accuracy (Díaz-Uriarte and De Andres, 2006), the accuracy might vary (Verikas et al., 2011). Thus, 1000 trees with 14 *mtrys* (i.e., using 1, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 24, 28, or 32 traits), were constructed in RF to determine the importance of each root trait in each pairwise comparison. In a third step, root traits were ranked in each pair based on their importance. Root traits importance was calculated in RF based on unbiased conditional inference permutation test (Strobl and Zeileis, 2008; Ball et al., 2014). The most important root trait was defined as

the one leading to the highest mean decrease of classification accuracy when values of a variable are randomly permuted across all 1000 trees (Breiman, 2001). Because negative importance values are due to random variation around zero (Strobl et al., 2009), values of root traits importance were first subtracted by the absolute value of the lowest negative importance and then normalized between 0 and 1 before being ranked in each pair.

Cultivar classification by either SVMs or RF were conducted through passing different combinations of top ranking important root traits derived from RF to SVMs/RF classification models. In order to find the combination of top ranking important root traits (*Timp*) that generate SVMs/RF models with the highest overall prediction accuracy, different numbers of *Timps* (*Timp_i*) were tested in SVMs/RF classification. Even though the number of variables in final SVMs models should generally be < 10 (Nicodemus et al., 2010), nine trait combinations were tested: top *i* important root traits (*Timp_i* = 2, 3, 5, 7, 9, 11, 13, 15) and all root traits (36). In rbf SVMs models, twelve kernel parameter *C* and regularization parameter *gamma* (Meyer, 2015) from 10^{-5} to 10^6 were tuned; the best combination of *C* and *gamma* leading to the highest prediction accuracy was chosen through leave-one-out cross-validation (LOOCV; Kohavi, 1995). Each SVMs pairwise classification model was validated with LOOCV. The accuracies of RF classification were derived from out of bag error (OOB); *Timp* = 1 was used to find the maximum HACCs when tuning different *mtrys* and *Timps* combinations. The validation accuracy of models means whether the different labeled observations were accurately classified.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (1)$$

Where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives. The validation accuracy was treated as final prediction accuracy of SVMs/RF classifications. Classifications with an average prediction accuracy $\geq 80\%$ were regarded as a high accuracy classifications (HACCs); the 80% level was determined acceptable by previous ML studies (Wang et al., 2010; Liu et al., 2014; Shang and Chisholm, 2014; Zheng et al., 2014; Sacchet et al., 2015). The whole process – RF ranking of root traits in each cultivar pair, SVMs and RF classification of pairs using different *mtrys* and *Timps* – was repeated three times; the average accuracy with standard error was calculated. The combination of different *mtrys* and different *Timp_i*s generating the highest average accuracy of SVM/RF models was treated as optimal *mtry* and *Timp_i* combination; the frequencies of the corresponding top five important root traits (T5IRT) from all HACCs were calculated. Because SVM models yielded higher classification accuracies than RF (**Figure 1**; Supplementary Figure S2), classification by RF was not pursued further. Subsequently the accuracies of SVMs models derived from *Timp₅* were compared to six runs of randomly selected subsets of five root traits each (R_5.1–R_5.6) to determine the benefits of root trait selection based on RF for cultivar classification.

The R for Windows package ‘party’ (v. 1.0–23) was used for RF trait ranking; the R package ‘e1071’ (v. 1.6–7) was used for rbf SVMs classification (data scaled), the ‘randomForest’

TABLE 2 | Abbreviation and description of 36 root traits derived either from direct measurement or from calculation after manual phenotyping of 16 European *P. sativum* cultivars.

Abbreviation	Unit	Definition	Level of trait
slatdiam2.5	mm	The average diameter of single lateral root originated from 0 to 5 cm deep tap root	Lateral roots
slatlth2.5	cm	The average length of single lateral root originated from 0 to 5 cm deep tap root	
slatsa2.5	cm ²	The average surface area of single lateral root originated from 0 to 5 cm deep tap root	
sLSRL2.5	cm g ⁻¹	The average specific root length (root length divided by dry biomass) of single lateral root originated from 0 to 5 cm deep tap root	
sLSRA2.5	cm ² g ⁻¹	The average specific root area (root area divided by dry biomass) of single lateral root originated from 0 to 5 cm deep tap root	
sLTD2.5	g cm ⁻³	The average tissue density (root dry biomass divided by root volume) of single lateral root originated from 0-5 cm deep tap root	
slatdw2.5	g	The average dry weight of single lateral root originated from 0 to 5 cm deep tap root	
latRL2.5	cm	The length of all the lateral roots originated from 0 to 5 cm deep tap root	
latRL7.5	cm	The length of all the lateral roots originated from 5 to 10 cm deep tap root	
latRSA2.5	cm ²	The surface area of all the lateral roots originated from 0 to 5 cm deep tap root	
latRSA7.5	cm ²	The surface area of all the lateral roots originated from 5 to 10 cm deep tap root	
latRV2.5	cm ³	The volume of all lateral roots coming originated from 0 to 5 cm deep tap root	
latRV7.5	cm ³	The volume of all lateral roots coming originated from 5 to 10 cm deep tap root	
latRDW2.5	g	All lateral roots dry weight from 0 to 5 cm deep tap root	
latRDW7.5	g	All lateral roots dry weight from 5 to 10 cm deep tap root	
latRDW	g	The dry weight of all lateral roots	Tap root
latn2.5		The number of lateral roots originated from 0 to 2.5 cm deep tap root	
latn5		The number of lateral roots originated from 2.5 to 5 cm deep tap root	
latn7.5		The number of lateral roots originated from 5 to 10 cm deep tap root	
tapdiam2.5	mm	The diameter of tap root from 0 to 5 cm depth	
tapdiam7.5	mm	The diameter of tap root from 5 to 10 cm depth	
tapTD2.5	g cm ⁻³	The tissue density of tap root from 0 to 5 cm depth	
tapTD7.5	g cm ⁻³	The tissue density of tap root from 5 to 10 cm depth	
tapdw2.5	g	Tap root dry weight from 0 to 5 cm depth	
tapdw7.5	g	Tap root dry weight from 5 to 10 cm depth	
tapRL	cm	Tap root length	
tapRSA	cm ²	Tap root surface area	
tapRV	cm ³	Tap root volume	
tapRDW	g	Tap root dry weight	Root system
lateralapRDWR	g g ⁻¹	The ratio of all lateral roots dry weight and tap root dry weight	
lateralapSAR	cm ² cm ⁻²	The ratio of all lateral roots area to tap root area	
totalRSA	cm ²	Total root surface area of whole root system	
totalRL	cm	Total root length of whole root system	
totalRV	cm ³	Total root volume of whole root system	
rootdw	g	Root dry weight	
rootdep	cm	Root depth	

package (v. 4.6–12) for RF classification. The R code used for RF root traits ranking and rbf SVMs classification can be found as Supplementary Method S3.

Univariate Permutation Test

In order to compare the efficiency between a univariate test and the combination of RF and rbf SVMs in cultivar classification, an exact permutation test (Hesterberg et al., 2005) was carried out ($\alpha = 0.05$), with both Bonferroni and *fdr* correction (Benjamini and Hochberg, 1995), to compare the root traits involved in each pairwise classification. The R code used can be found as Supplementary Method S4.

RESULTS

RF Root Traits Selection and SVMs Classification

The classification accuracy of SVMs multiclass classifications with and without root trait selection were 16.5 and 22.7%, respectively (Supplementary Figure S1), thus a pairwise approach was pursued thereafter. Similarly, the number of HACCs generated from different *mtrys* and *Timpis* combinations in pairwise RF (77 HACCs with *mtry* = 32, *Timpis* = 2; Supplementary Figure S2) were much lower compared to the classification accuracy achieved by pairwise SVMs (see below and **Figure 1**). Thus, this RF/RF approach was not followed

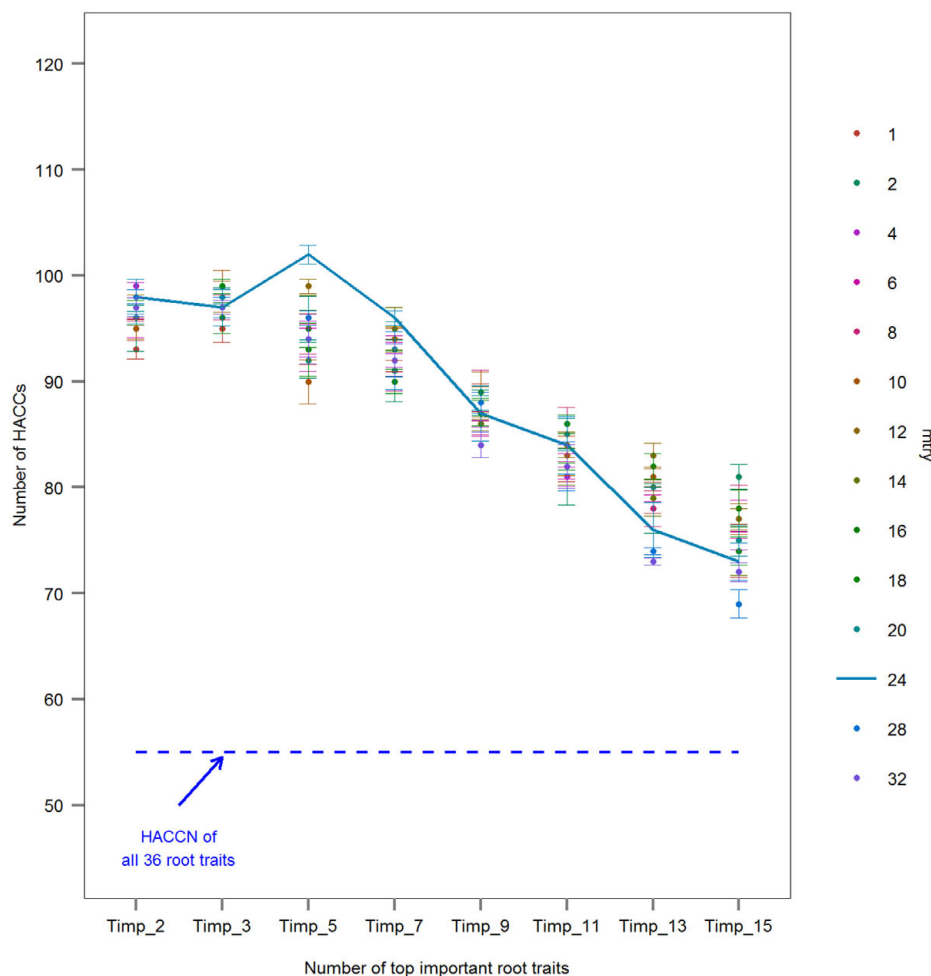


FIGURE 1 | Comparison of the number of high accuracy classifications (HACCs; prediction accuracy $\geq 80\%$) using rbf SVMs models with different numbers of *Pisum sativum* root traits randomly chosen for each tree, *mtrys*, and different number of top important root traits, *Timp_i* = 2, 3, 5, 7, 9, 11, 13, 15 (mean \pm SE; $n = 3$ runs). One hundred twenty pairs of 16 European *P. sativum* cultivars were analyzed. A dashed blue line represents the number of HACCs (HACCN) retrieved using all 36 root traits in SVMs classification. A line connecting HACC numbers derived by using different number of *Timps* at constant *mtry* = 24 was added for visualization.

further but a combination of RF (for trait ranking) and SVMs (for classification) was used. Through testing a series of combinations of different *mtrys* and *Timpis*, *mtry* = 24 and *Timp_i* = 5 (Timp5) generated the highest number of HACCs in pairwise comparison, 101 (averaged from three runs; 100, 101, 103 HACCs) out of 120 pairs (**Figure 1**). Although other combinations of *mtrys* and single-digit *Timpis* resulted in similar numbers of HACCs, the run which contained the highest number of HACCs (103) was used exemplary for further analysis (**Figure 2**). The least number of HACCs (54) was computed from SVMs models using all available root traits (36). The number of HACCs thus increased by 91% through root trait selection with RF and tuning SVMs to Timp5. Both single SVMs model accuracy and the total number of HACCs based on six runs of five randomly selected root traits each (R_5.1–5.6) are similar or decreased compared to using all 36 root traits (All_36) and much lower than SVMs models

involving the top five important root traits (Timp_5), respectively (**Figure 3**).

The Timp5 in each of the 103 HACCs are ranked as 1st, 2nd, 3rd, 4th, and 5th and indicated with different colors in **Figure 4** (see Supplementary Table S1 for a list of Timp5 and Supplementary Table S2 for normalized importance values of Timp5). The most frequent Timp5 of the analyzed pea cultivars in all HACCs (T5IRT) are latRSA2.5, tapdiam2.5, latn2.5, tapdw7.5, and totalRSA (see **Table 2** for trait abbreviations) with proportions of 31, 23, 21, 21, and 21%, respectively (**Figure 5**). The root traits measured at both 0–5 and 5–10 cm depth (along the tap root) are lateral root surface area, tap root diameter, lateral root number, tap root dry weight (tapRDW), lateral root length, lateral root dry weight (latRDW), lateral root volume, and tap root TD. Four out of eight root trait pairs measured at 0–5 cm depth (i.e., latRSA2.5, tapdiam2.5, latn2.5, latn5, and tapTD2.5)

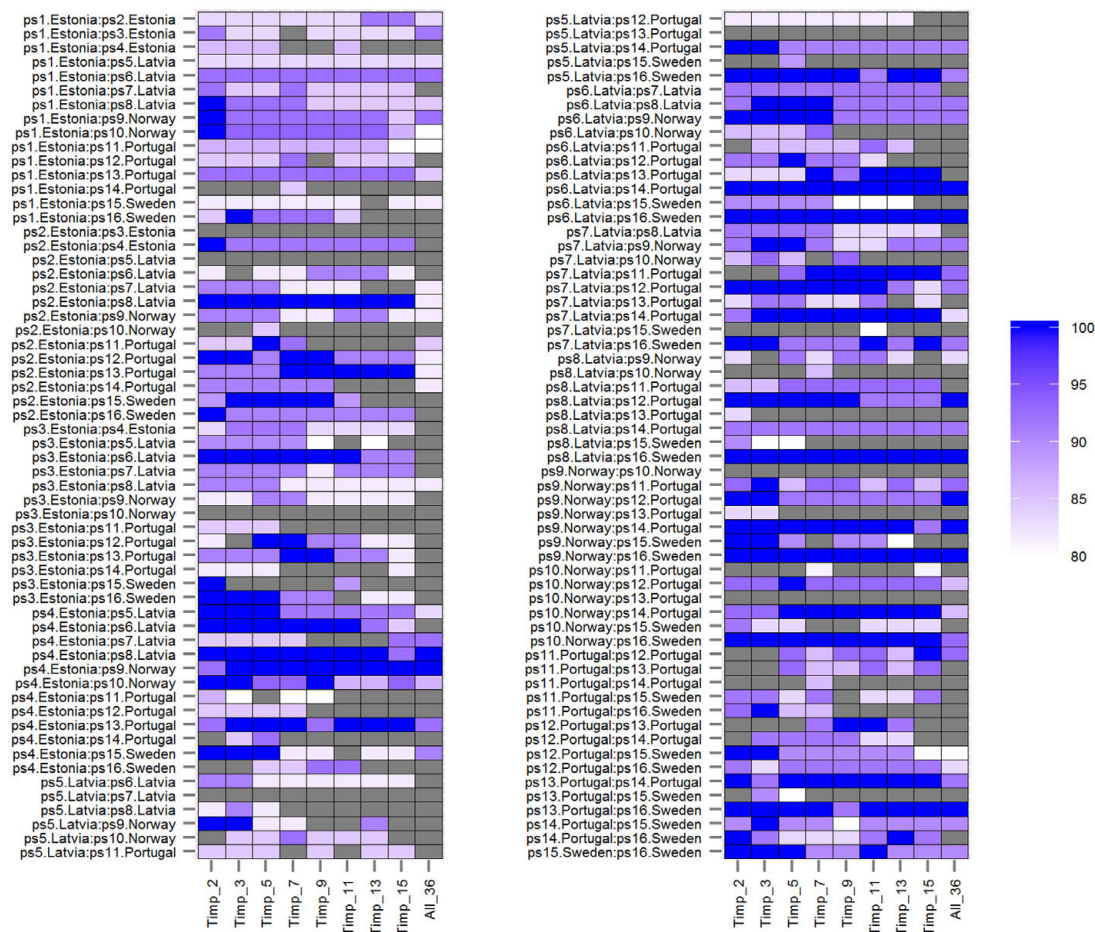


FIGURE 2 | Prediction accuracies of rbf SVMs models ($mtry = 24$) using different numbers of top important root traits ($Timp_j = 2, 3, 5, 7, 9, 11, 13, 15$) and all (36) root traits (All_36) for 120 pairwise comparisons among 16 European *P. sativum* cultivars. Prediction accuracies from 80% (white) to 100% (blue) are shown as color gradient; lower accuracies are displayed in gray. Cultivars are listed in Table 1.

have a higher frequency among all Timp5 than corresponding ones from 5–10 cm depth along the tap root while the other four have similar frequencies (Figure 5, inset); the average frequency of Timp5 from 0–5 cm depth is 18 compared to 14 from 5–10 cm depth along the tap root.

In order to confirm whether conditional inference permutation test can rank root traits without bias, correlation coefficients of Timp5s are compared in different pairwise classifications. For example, Timp5 in the pair ps1.Estonia vs. ps2.Estonia (Figure 4, first line; Supplementary Tables S1 and S2) are tapRDW between 5–10 cm depth (tapdw7.5), lateral root try weight of 0–5 cm depth (latRDW2.5), total root system dry weight (rootdw), latRDW, and SRL of lateral roots at 0–5 cm depth (sSRL2.5). In this pair, the highest Pearson correlation between the most important root trait tapdw7.5 (ranked 1st) and the other four traits is 0.52 (Figure 6; see Supplementary Figure S3 for Spearman correlation); tapRDW, which has a very high correlation coefficient of 0.89 with tapdw7.5, is not involved in the Timp5 of the pair ps1.Estonia vs. ps2.Estonia. Rootdw and latRDW, which are highly correlated with a Pearson coefficient

of 1, ranked as 3rd and 4th in the classification of ps1.Estonia and ps2.Estonia (Figure 6). In another example (ps1.Estonia vs. ps7.Latvia), both rootdw and latRDW are among the Timp5 but rootdw is ranked 1st while latRDW is ranked 5th (Figure 4; Supplementary Tables S1 and S2).

The highest correlation coefficient among T5IRTs was 0.93 (Pearson correlation, Figure 6; see Supplementary Figure S3 for Spearman correlations) between latRSA2.5 and totalRSA while the lowest was 0.31 (Pearson) between tapdw7.5 and lateral root number at 0–2.5 cm depth (latn2.5). The frequencies of latRSA2.5 and totalRSA among T5IRTs (Figure 5) are among the highest with 32 and 22, respectively, while they only appear nine times simultaneously in the same pairwise classification.

Permutation Test Comparing the Difference of the Mean of Single Root Trait

Comparing the efficiency of ML techniques conducted by RF and SVMs with a univariate permutation test, only 46.6%

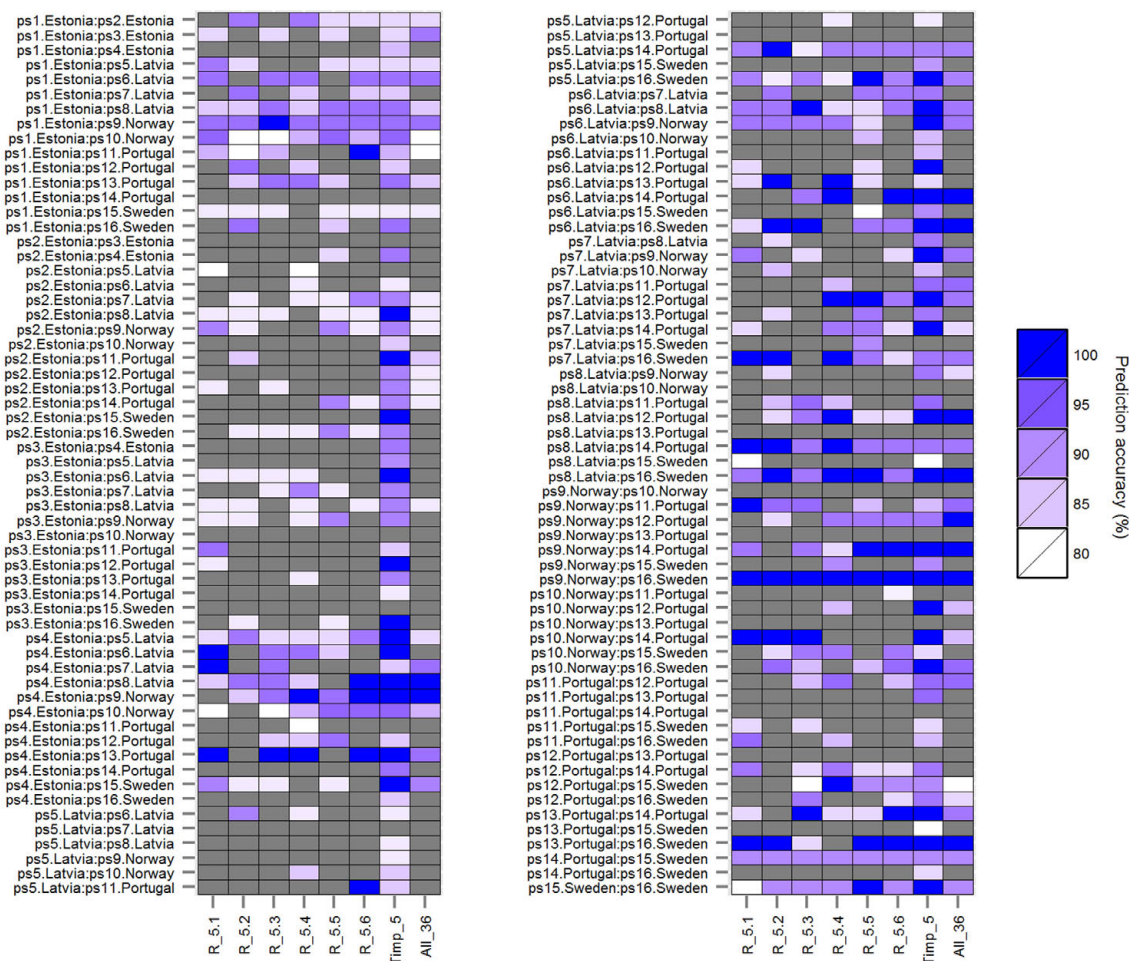


FIGURE 3 | Prediction accuracies of rbf SVMs models ($mtry = 24$) using five randomly selected root traits each in six runs ($R_{5.1}$, $R_{5.2}$, $R_{5.3}$, $R_{5.4}$, $R_{5.5}$, and $R_{5.6}$) compared to prediction accuracies using top five important root traits ($Timp_5$), and all (36) root traits (All_36) for 120 pairwise comparisons among 16 European *P. sativum* cultivars. Prediction accuracies from 80% (white) to 100% (blue) are shown as color gradient; lower accuracies are displayed in gray. Cultivars are listed in Table 1.

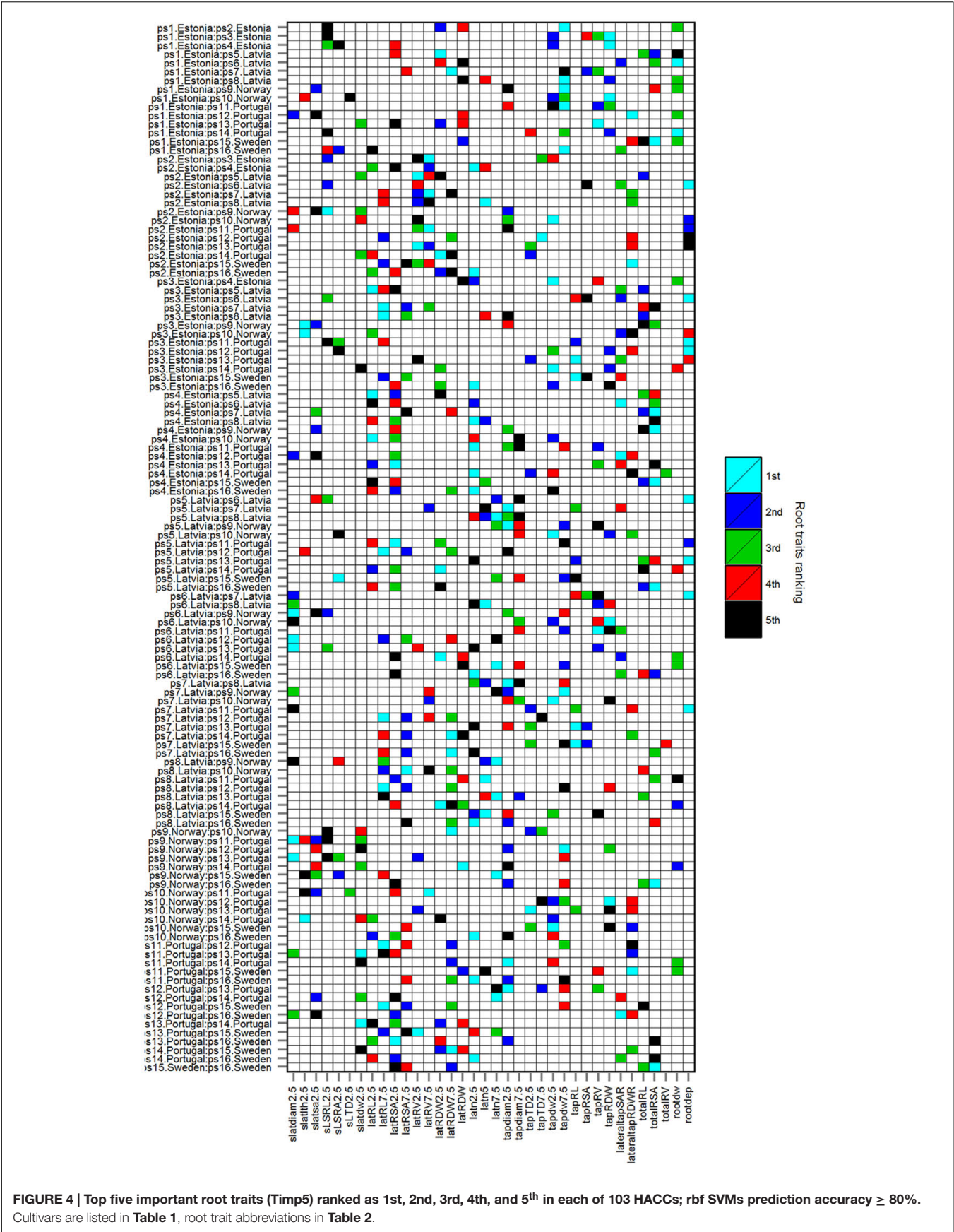
(Bonferroni-corrected) and 47.5% (fdr-corrected) of HACCs (based on $Timp5$) have significantly different root traits. Significantly different root traits in univariate permutation tests can be found in (supplementary) figures (Figure 7, with Bonferroni correction; Supplementary Figure S4, with fdr correction). SVMs were not always superior to univariate permutation test without root traits selection: classification accuracies of several pairwise SVMs classifications involving significantly different root traits in univariate permutation test were lower than 80% [see, e.g., ps11.Portugal vs. ps16.Sweden (Figure 7; Supplementary Figure S4)].

Three examples are given to visualize the mean difference of root traits in pairwise comparisons. There are significant different root traits between ps1.Estonia vs. ps8.Latvia and the cultivars are also classified with high accuracy (92.3%; Figure 8). In contrast, cultivars ps3.Estonia vs. ps9.Norway are classified with high accuracy (90.9%), however, without any significant root traits identified by univariate permutation test (Supplementary

Figure S5). There are (visibly) no significant different root traits between cultivars ps1.Estonia vs. ps14.Portugal and they are not accurately classified either – accuracy 69.2% (Supplementary Figure S6). The most significantly different root traits (the lowest p value in univariate permutation test) can be different from the most important root traits (ranked by RF) in each pairwise classification, e.g., $Timp5$ in the pair ps3.Estonia vs. ps9.Norway are tapdw7.5, tapRDW, rootdw, latn5, and latRDW while the rank order based on p values from permutation test changed to latn5, tapdw7.5, tapRDW, latRDW, and rootdw. No tied importance values of $Timp5$, indicating higher ranking root traits are more important in RF, have been found.

DISCUSSION

Machine learning algorithms are promising statistical tools to assist humans in the analysis of complex data sets and



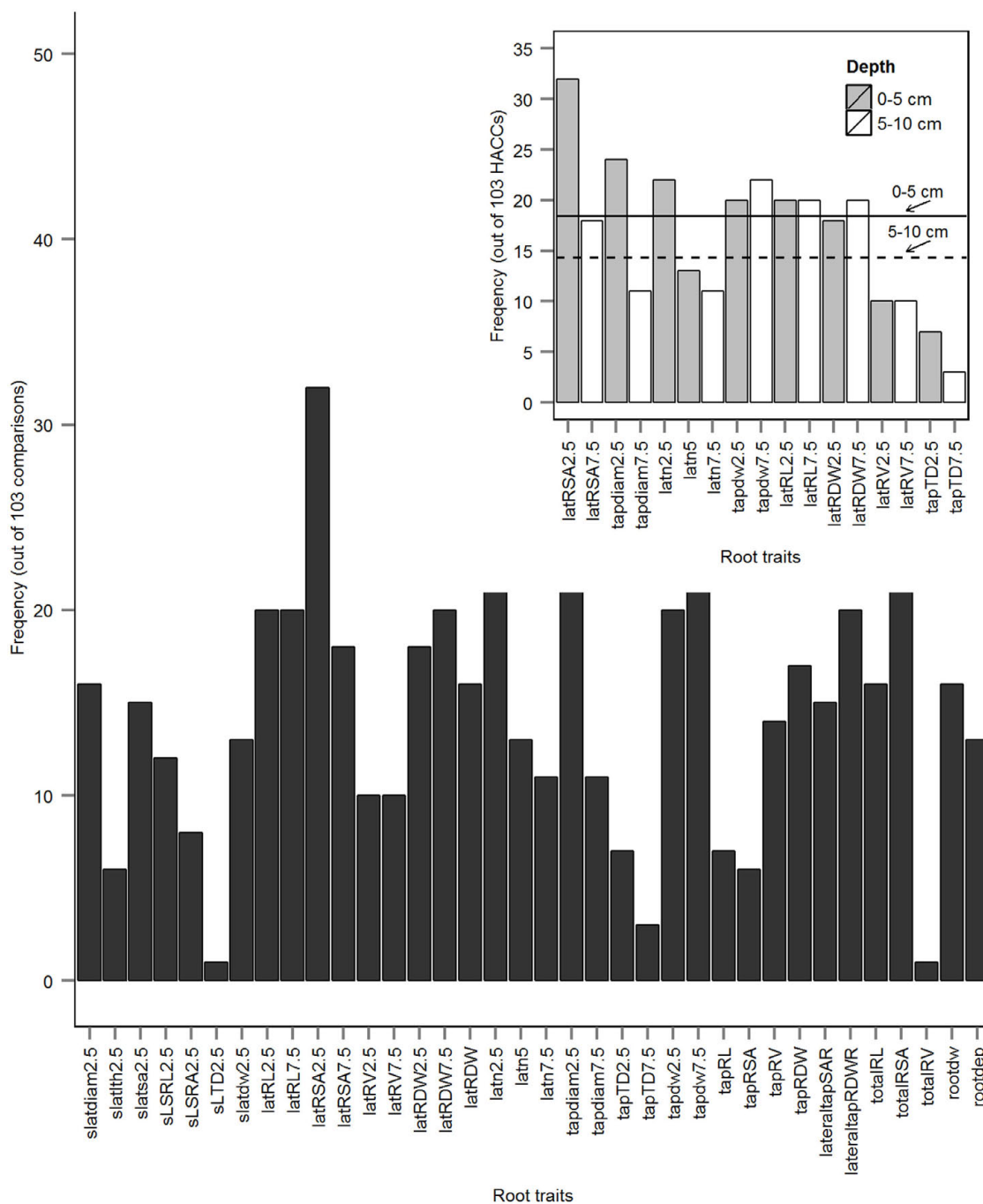
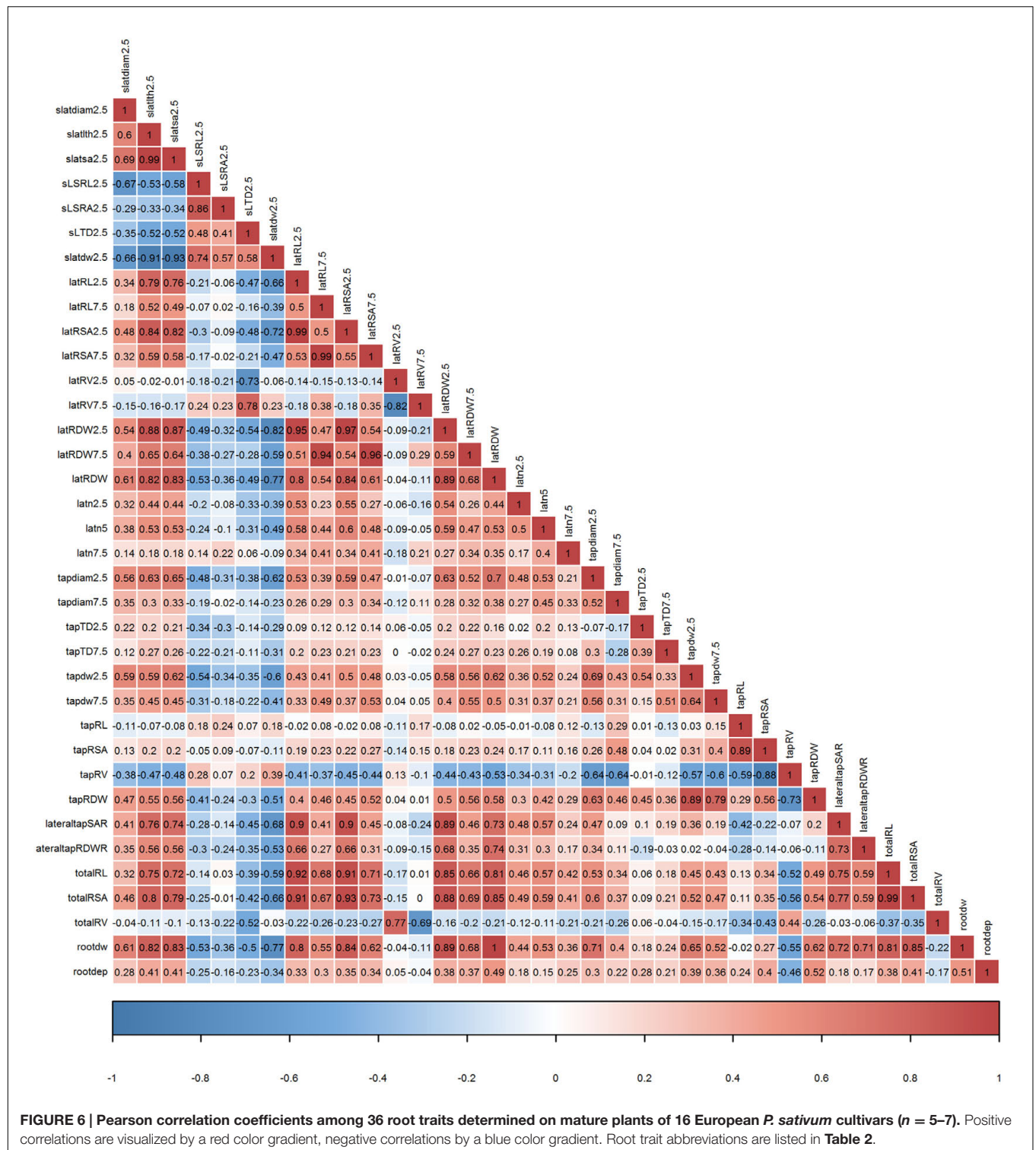


FIGURE 5 | Frequencies of top five important root traits (Timp5) in 103 high accuracy classifications (HACCs; rbf SVMs prediction accuracy $\geq 80\%$) identified out of 120 pairwise comparisons among 16 European *P. sativum* cultivars. The inset illustrates the frequencies of root traits from 0 to 5 cm depth along the tap root (gray bars) and from 5 to 10 cm depth (white bars), respectively; a solid line indicates the average number of Timp5 from 0 to 5 cm depth, a dashed line indicates the average number of Timp5 from 5 to 10 cm depth tap root. Root trait abbreviations are listed in **Table 2**.

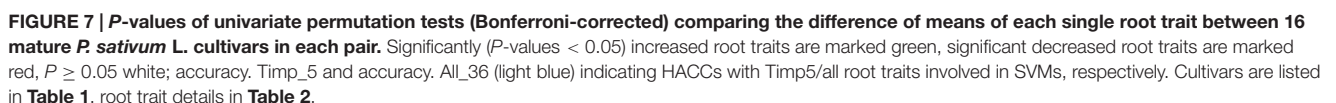
started to be widely used in many research fields including (plant) genomics/proteomics (Ma et al., 2014; Libbrecht and Noble, 2015). To the best of our knowledge, they have been only applied twice on RSA differentiation yet (Zhong et al., 2009; Iyer-Pascuzzi et al., 2010). This is surprising because

as new technologies for generating large plant phenotypic data sets emerge, demand will drastically increase for new statistical techniques. Phenotyping is estimated in becoming the major operational bottleneck in limiting the power of genetic analysis and genomic prediction (Rahaman et al., 2015). Data



complexity of particular sets of traits is generally high, especially in root systems where the developments causing a specific architecture, and the physiology and performance of individual root segments/units within the branched root system are not well understood yet.

So far, most of the techniques developed for RSA phenotyping involve the use of seedlings (Kuijken et al., 2015). Although there are examples in which the early stage root phenotype has predictive value for later developmental stages (Tuberosa et al., 2002), the seedling root phenotype may not always be



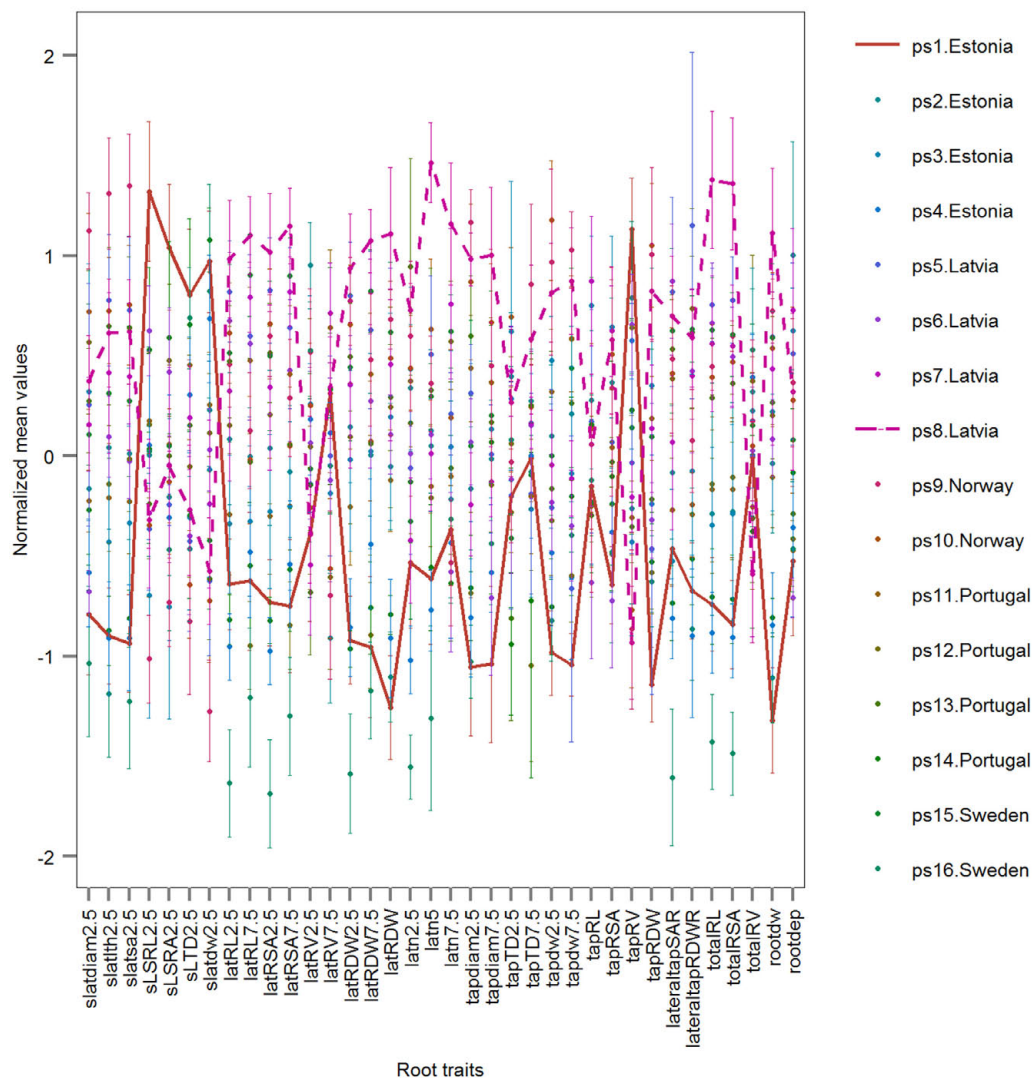


FIGURE 8 | Normalized root traits determined on mature plants of 16 European *P. sativum* cultivars (mean \pm SE; $n = 5-7$). Two horizontal lines visualize traits of the cultivars ps1.Estonia (solid) and ps8.Latvia (dashed). Cultivars are listed in **Table 1**, root trait abbreviations in **Table 2**.

representative of the mature plant (Watt et al., 2013). Because replicate numbers from manual phenotyping of mature root systems are limited but mature root systems often have a higher complexity, adapted statistical methods need to be developed to make full use of data sets. Here we applied combined ML algorithms with unbiased variable importance measure for the first time successfully, to the best of our knowledge, on a small RSA/root morphology data set manually derived from 97 mature plants of 16 European pea cultivars. The importance of 36 root traits was measured and ranked in RF. Pairwise classifications were analyzed either by SVMs based on a Gaussian radial basis kernel function (rbf SVMs) or by RF (standard algorithm) with the RF-identified top ranking root traits. The overall accuracy of models was cross-validated. The combination of SVMs and RF improved the classification accuracy – confirming earlier results by Löw et al. (2012) in remote sensing.

When compared to classical statistical tools, our results demonstrated that all pairwise classifications with significant root traits from univariate permutation test belong to HACCs with Timp5, however, almost half of HACCs derived from Timp5 don't have any significantly different root traits. This points to the advantages of combining RF and SVMs in root traits importance measure and cultivar classification. Besides robustness to noises, RF considers both the influence of single variables separately and the multivariate interactions with other variables, which make this advanced ML approach more efficient, accurate, and reliable (Breiman, 2001; Zhu et al., 2012). Among the HACCs with significantly different root traits, the ranking of top five important root traits (Timp5) was not matched by p values from univariate permutation test, i.e., the most significantly different root traits differ with the most important root traits identified by pairwise classification. This might be due to that SVMs classification

concerns more about parameters (root traits) involved in the SVMs models while traditional multi/univariate analyses focus more on differences of specific root traits between two groups (Gaonkar et al., 2015). We conclude that the combination of RF root traits selection and SVMs classification can make full use of all possible information of root traits in pea cultivars' classification.

Our results clearly demonstrated the importance of selecting important root traits by RF to obtain an efficient classification based on RSA among dicot crop cultivars. SVMs models using all available traits or including five randomly selected root traits (R_5) were not able to increase the overall accuracy, which confirmed the necessity of root traits selection through RF in cultivar differentiation. This finding is in accordance with previous ML approaches in other scientific fields (Wang et al., 2010; Löw et al., 2012; Liu et al., 2014). The improved accuracy probably benefits from alleviating the 'curse of dimensionality' through root traits selection, removing non-informative signals (Chu et al., 2012). Thus, we can show that the identification of a few important root traits, in our case five, significantly increases the classification accuracy. While we did not find any single pea root trait that was always more important than others in all HACCs, a more targeted cultivar differentiation and trait selection for breeding can be obtained when focusing on root traits with highest frequency among T5IRTs. The most frequent T5IRT among the tested pea cultivars was the surface area of all lateral roots originating at the tap root between 0–5 cm (latRSA2.5) – appearing in more than one third of the pairwise comparisons. Distinguishing cultivars based on latRSA2.5 value can have important ecological effects: greater latRSA2.5 implicates more absorptive lateral root surface area in the topsoil and thus the potential for enhanced P or topsoil water foraging. Lower latRSA2.5 values means that plants are possibly privileging deep soil exploration (Miguel et al., 2015) with potential influences on drought tolerance or performance in low input agriculture (Bonser et al., 1996). Another frequent T5IRT in pea was latn2.5, the number of laterals originating at the tap root between 0–5 cm, which is somehow comparable to the trait 'whorl numbers' of *Phaseolus vulgaris* seedlings (Miguel et al., 2013). Miguel and colleagues could show that common bean genotypes with greater whorl numbers accumulated up to 60% more biomass under low-phosphorus conditions.

Completely intact root systems can hardly be collected by destructive harvesting methods, especially of mature plants with deeper root systems grown in the field; similar, lateral root traits are likely more affected by destructive sampling, e.g., by root tip shedding, than tap root traits. RSA information retrieved from top soil layers is thus likely more accurate (Miguel, 2012). Interestingly, the number of Timp5 derived from 0–5 cm depth of the pea tap root were more frequent than the ones originating from 5–10 cm depth, indicating that root traits from the top of the tap root have a greater potential to differentiate pea cultivars. This knowledge is already utilized by 'shovelomics' approaches, which only excavate the root crown of mature plants for phenotypical analysis (Trachsel et al., 2011; Bucksch et al., 2014). Our findings thus provide additional evidence that shovelomics can be considered an

informative field-based high-throughput phenotyping approach due to the strong contribution of topsoil root traits to cultivar distinction.

Root system depth and average radius were previously identified as frequently top-ranked root traits in linear SVMs classification to distinguish different rice genotypes (Iyer-Pascuzzi et al., 2010). In our study the top two frequently important root traits were latRSA2.5 and tapdiam2.5 while root system depth (rootdep) was much less frequently present in Timp5. However, the top important root traits identified by Iyer-Pascuzzi et al. (2010) might be subject to change as the used ranking method was recently deemed biased (Gaonkar and Davatzikos, 2013). However, the difference of key root traits is likely also related to species-specific differences between rice and pea, but also the differences in growth stages (juvenile vs. mature), media (gel vs. sand), and analyzing methods (see "Discussion" above).

Correlation among traits is generally considered as an indication of their redundancy for classification. However, they may still provide complementary information and an otherwise inconclusive variable can provide a significant performance in combination with others (Guyon and Elisseeff, 2003). The correlation of Timp5 from all pairwise classifications varied greatly in this study: On the one hand, highly correlated root traits were not always top-ranked; on the other hand, root traits that were highly correlated with the most important Timp could even not be important at all in our study. The correlation variance of Timp5 thus confirms the unbiased root traits importance measure through conditional inference permutation test – increasing data interpretability.

CONCLUSION

The accurate classification of 86% (103 of 120) genotype pairs of pea indicated that most of the studied cultivars could be well differentiated by using a few most distinguishing root traits, as selected through RF. This implies that past culturing did not lead to a major loss of RSA variability of the studied European pea cultivars. Breeders are envisioned to work more effectively in future breeding programs by knowing distinguishing (pea) root traits in advance (Manavalan et al., 2010). In specific, pairwise classification approaches can help breeders to make informed decisions on cultivars selection for crossing. Powerful statistical approaches are essential to make use of the increasing amount of phenotyping information available, integrating the complex trait sets. In particular, this study showed that combining RF with rbf SVMs for variable selection and group classification, respectively, can overcome problems in applying ML approaches on data sets characterized by a rather low signal-to-noise ratio. Thus, ML methods are generally envisioned to make plant phenotypical data analyses more effectively, robust and comprehensive. However, our experiment under standardized conditions might have caused the loss of root traits adaptive to local environmental conditions. Thus, further ML-supported analysis of field-derived root phenotypes under varying environments are urgently needed, selecting genotypes that feature specific sets of traits

facilitating plant performance under local edaphic and climatic conditions. Advanced methods must be urgently developed in order to facilitate the phenotyping of mature root systems under realistic growing conditions.

AUTHOR CONTRIBUTIONS

JZ, GB, and BR conceived and planned the experiment; JZ and BR performed the experiment; JZ analyzed the data; all authors jointly wrote the manuscript.

FUNDING

This project received funding from the European Union's Seventh Framework Program for research, technological

development and demonstration under grant agreement no 613781 (EUROLEGUME).

ACKNOWLEDGMENTS

We kindly thank all donors for the genetic material; Ina Alsina, Latvia and Aleš Látr, Czech Republic kindly proved the rhizobia and mycorrhizal inoculum, respectively. Michael J. Bambricks' help during the harvest and root system dissection was essential.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01864/full#supplementary-material>

REFERENCES

- Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A. L., Jacquin, F., et al. (2015). Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 84, 1–19. doi: 10.1111/tpj.12967
- Annicchiarico, P., and Iannucci, A. (2008). Breeding strategy for faba bean in southern europe based on cultivar responses across climatically contrasting environments. *Crop Sci.* 48, 983–991. doi: 10.2135/cropsci2007.09.0501
- Ball, T. M., Stein, M. B., Ramsawh, H. J., Campbell-Sills, L., and Paulus, M. P. (2014). Single-subject anxiety treatment outcome prediction using functional neuroimaging. *Neuropsychopharmacology* 39, 1254–1261. doi: 10.1038/npp.2013.328
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.
- Berger, B., Parent, B., and Tester, M. (2010). High-throughput shoot imaging to study drought responses. *J. Exp. Bot.* 61, 3519–3528. doi: 10.1093/jxb/erq201
- Bonser, A. M., Lynch, J., and Snapp, S. (1996). Effect of phosphorus deficiency on growth angle of basal roots in *Phaseolus vulgaris*. *New Phytol.* 132, 281–288. doi: 10.1111/j.1469-8137.1996.tb01847.x
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1017934522171
- Bucksch, A., Burridge, J., York, L. M., Das, A., Nord, E., Weitz, J. S., et al. (2014). Image-based high-throughput field phenotyping of crop roots. *Plant Physiol.* 166, 470–486. doi: 10.1104/pp.114.243519
- Caringella, M. A., Bongers, F. J., and Sack, L. (2015). Leaf hydraulic conductance varies with vein anatomy across *Arabidopsis thaliana* wild-type and leaf vein mutants. *Plant Cell Environ.* 38, 2735–2735. doi: 10.1111/pce.12584
- Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M., and Makowski, D. (2015). Estimating variability in grain legume yields across Europe and the Americas. *Sci. Rep.* 5:11171. doi: 10.1038/srep11171
- Chu, C., Hsu, A.-L., Chou, K.-H., Bandettini, P., Lin, C., Initiative, et al. (2012). Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 60, 59–70. doi: 10.1016/j.neuroimage.2011.11.066
- Cortes, C., and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.* 20, 273–297. doi: 10.1007/BF00994018
- Cramer, G. R., Ergül, A., Grimplet, J., Tillet, R. L., Tattersall, E. A., Bohlman, M. C., et al. (2007). Water and salinity stress in grapevines: early and late changes in transcript and metabolite profiles. *Funct. Integr. Genomics* 7, 111–134. doi: 10.1007/s10142-006-0039-y
- da Costa, L. L., Delcroix, M., Dalla Costa, E. R., Prestes, I. V., Milano, M., Francis, S. S., et al. (2015). A real-time PCR signature to discriminate between tuberculosis and other pulmonary diseases. *Tuberculosis* 95, 421–425. doi: 10.1016/j.tube.2015.04.008
- Díaz-Uriarte, R., and De Andres, S. A. (2006). Gene selection and classification of microarray data using random forest. *BMC Bioinformatics* 7:3. doi: 10.1186/1471-2105-7-3
- EU (2013). *The Environmental Role of Protein Crops in the New Common Agricultural Policy: Agricultural and Rural Development: Directorate General for Internal Policies, Policy Department B: Structural and Cohesion Policies*. Brussels: European Parliament, 119.
- FAOSTAT (2014). *Food and Agriculture Organization of the United Nations*. Available at: <http://faostat3.fao.org/home/E>
- Fiorani, F., and Schurr, U. (2013). Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* 64, 267–291. doi: 10.1146/annurev-arplant-050312-120137
- Gaonkar, B., and Davatzikos, C. (2013). Analytic estimation of statistical significance maps for support vector machine based multi-variate image analysis and classification. *Neuroimage* 78, 270–283. doi: 10.1016/j.neuroimage.2013.03.066
- Gaonkar, B., Shinohara, R. T., Davatzikos, C., Alzheimers Disease, and Neuroimaging Initiative. (2015). Interpreting support vector machine models for multivariate group wise analysis in neuroimaging. *Med. Image Anal.* 24, 190–204. doi: 10.1016/j.media.2015.06.008
- Gilbert, M. E., Zwieniecki, M. A., and Holbrook, N. M. (2011). Independent variation in photosynthetic capacity and stomatal conductance leads to differences in intrinsic water use efficiency in 11 soybean genotypes before and during mild drought. *J. Exp. Bot.* 62, 2875–2887. doi: 10.1093/jxb/erq461
- Goebel, E., and Calatayud, A. (2012). Applications of chlorophyll fluorescence imaging technique in horticultural research: a review. *Sci. Hortic.* 138, 24–35. doi: 10.1016/j.scienta.2012.02.002
- Gowin, J. L., Ball, T. M., Wittmann, M., Tapert, S. F., and Paulus, M. P. (2015). Individualized relapse prediction: personality measures and striatal and insular activity during reward-processing robustly predict relapse. *Drug Alcohol Depend.* 152, 93–101. doi: 10.1016/j.drugalcdep.2015.04.018
- Grady, K. C., Laughlin, D. C., Ferrier, S. M., Kolb, T. E., Hart, S. C., Allan, G. J., et al. (2013). Conservative leaf economic traits correlate with fast growth of genotypes of a foundation riparian species near the thermal maximum extent of its geographic range. *Funct. Ecol.* 27, 428–438. doi: 10.1111/1365-2435.12060
- Guo, D., Xia, M., Wei, X., Chang, W., Liu, Y., and Wang, Z. (2008). Anatomical traits associated with absorption and mycorrhizal colonization are linked to root branch order in twenty-three Chinese temperate tree species. *New Phytol.* 180, 673–683. doi: 10.1111/j.1469-8137.2008.02573.x
- Guyon, I., and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.* 3, 1157–1182.
- Hargreaves, C. E., Gregory, P. J., and Bengough, A. G. (2009). Measuring root traits in barley (*Hordeum vulgare* ssp. *vulgare* and ssp. *spontaneum*) seedlings using gel chambers, soil sacs and X-ray microtomography. *Plant Soil* 316, 285–297. doi: 10.1007/s11104-008-9780-4

- Hartmann, A., Czauderna, T., Hoffmann, R., Stein, N., and Schreiber, F. (2011). HTPPheno: an image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics* 12:148. doi: 10.1186/1471-2105-12-148
- Hesterberg, T., Moore, D. S., Monaghan, S., Clipson, A., and Epstein, R. (2005). Bootstrap methods and permutation tests. *Introduction Pract. Stat.* 5, 1–70.
- Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2010). *A Practical Guide to Support Vector Classification*. Taipei: National Taiwan University.
- Hu, X., Tanaka, A., and Tanaka, R. (2013). Simple extraction methods that prevent the artifactual conversion of chlorophyll to chlorophyllide during pigment isolation from leaf samples. *Plant Methods* 9:19. doi: 10.1186/1746-4811-9-19
- Hund, A., Reimer, R., and Messmer, R. (2011). A consensus map of QTLs controlling the root length of maize. *Plant Soil* 344, 143–158. doi: 10.1007/s11104-011-0735-9
- Iqbal, Y., and Lewandowski, I. (2014). Inter-annual variation in biomass combustion quality traits over five years in fifteen *Miscanthus* genotypes in south Germany. *Fuel Process. Technol.* 121, 47–55. doi: 10.1016/j.fuproc.2014.01.003
- Iyer-Pascuzzi, A. S., Symonova, O., Mileyko, Y., Hao, Y., Belcher, H., Harer, J., et al. (2010). Imaging and analysis platform for automatic phenotyping and trait ranking of plant root systems. *Plant Physiol.* 152, 1148–1157. doi: 10.1104/pp.109.150748
- Jezierny, D., Mosenthin, R., and Bauer, E. (2010). The use of grain legumes as a protein source in pig nutrition: a review. *Anim. Feed Sci. Technol.* 157, 111–128. doi: 10.1016/j.anifeeds.2010.03.001
- Jia, Y., Gray, V. M., and Straker, C. J. (2004). The influence of *Rhizobium* and arbuscular mycorrhizal fungi on nitrogen and phosphorus accumulation by *Vicia faba*. *Ann. Bot.* 94, 251–258. doi: 10.1093/aob/mch135
- Jiang, H., Deng, Y., Chen, H.-S., Tao, L., Sha, Q., Chen, J., et al. (2004). Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics* 5:81. doi: 10.1186/1471-2105-5-81
- Kashiwagi, J., Krishnamurthy, L., Crouch, J., and Serraj, R. (2006). Variability of root length density and its contributions to seed yield in chickpea (*Cicer arietinum* L.) under terminal drought stress. *Field Crops Res.* 95, 171–181. doi: 10.1016/j.fcr.2005.02.012
- Kashiwagi, J., Krishnamurthy, L., Upadhyaya, H. D., Krishna, H., Chandra, S., Vadez, V., et al. (2005). Genetic variability of drought-avoidance root traits in the mini-core germplasm collection of chickpea (*Cicer arietinum* L.). *Euphytica* 146, 213–222. doi: 10.1007/s10681-005-9007-1
- Kohavi, R. (1995). “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *IJCAI’95 – Proceedings of the 14th international joint conference on Artificial intelligence*, Vol. 2, (San Francisco, CA: Morgan Kaufmann Publishers), 1137–1145.
- Koivunen, E., Tuunainen, P., Rossow, L., and Valaja, J. (2014). Digestibility and utilization of faba bean (*Vicia faba* L.) diets in broilers. *Acta Agric. Scand. Sect. A Anim. Sci.* 64, 217–225. doi: 10.1080/09064702.2015.1020857
- Kuijken, R. C. P., van Eeuwijk, F. A., Marcelis, L. F. M., and Bouwmeester, H. J. (2015). Root phenotyping: from component trait in the lab to breeding. *J. Exp. Bot.* 66, 5389–5401. doi: 10.1093/jxb/erv239
- Li, L., Zhang, Q., and Huang, D. (2014). A review of imaging techniques for plant phenotyping. *Sensors* 14, 20078–20111. doi: 10.3390/s141120078
- Libbrecht, M. W., and Noble, W. S. (2015). Machine learning applications in genetics and genomics. *Nat. Rev. Genet.* 16, 321–332. doi: 10.1038/nrg.3920
- Liu, N., Koh, Z. X., Goh, J., Lin, Z., Haaland, B., Ting, B. P., et al. (2014). Prediction of adverse cardiac events in emergency department patients with chest pain using machine learning for variable selection. *BMC Med. Inform. Decis. Mak.* 14:75. doi: 10.1186/1472-6947-14-75
- Liu, W., Gopal, S., and Woodcock, C. E. (2004). Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogramm. Eng. Remote Sensing* 70, 963–971. doi: 10.14358/PERS.70.8.963
- Lizarazo, C. I., Lampi, A. M., Liu, J., Sontag-Strohm, T., Piironen, V., and Stoddard, F. L. (2015). Nutritive quality and protein production from grain legumes in a boreal climate. *J. Sci. Food Agric.* 95, 2051–2064. doi: 10.1002/jsfa.6920
- Löw, F., Schorcht, G., Michel, U., Dech, S., and Conrad, C. (2012). “Per-field crop classification in irrigated agricultural regions in middle Asia using random forest and support vector machine ensemble,” in *Proceedings of the SPIE 8538, Earth Resources and Environmental Remote Sensing/GIS Applications III, 85380R*, (Edinburgh: International Society for Optics and Photonics), 1–11.
- Lynch, J. P. (2007). Roots of the second green revolution. *Aust. J. Bot.* 55, 493–512. doi: 10.1071/BT06118
- Lynch, J. P. (2011). Root phenes for enhanced soil exploration and phosphorus acquisition: tools for future crops. *Plant Physiol.* 156, 1041–1049. doi: 10.1104/pp.111.175414
- Ma, C., Zhang, H. H., and Wang, X. (2014). Machine learning for Big Data analytics in plants. *Trends Plant Sci.* 19, 798–808. doi: 10.1016/j.tplants.2014.08.004
- Mairhofer, S., Zappala, S., Tracy, S. R., Sturrock, C., Bennett, M., Mooney, S. J., et al. (2012). RooTrak: automated recovery of three-dimensional plant root architecture in soil from x-ray microcomputed tomography images using visual tracking. *Plant Physiol.* 158, 561–569. doi: 10.1104/pp.111.186221
- Manavalan, L. P., Guttikonda, S. K., Nguyen, V. T., Shannon, J. G., and Nguyen, H. T. (2010). Evaluation of diverse soybean germplasm for root growth and architecture. *Plant Soil* 330, 503–514. doi: 10.1007/s11104-009-0222-8
- Meyer, D. (2015). *Support Vector Machines. The Interface to Libsvm in Package e1071*. Vienna: FH Technikum Wien.
- Miguel, M. A. (2012). *Functional Role and Synergistic Effect of Root Traits for Phosphorus Acquisition Efficiency and Their Genetic Basis in Common Bean (Phaseolus vulgaris L.)*. State College, PA: The Pennsylvania State University.
- Miguel, M. A., Postma, J. A., and Lynch, J. P. (2015). Phene synergism between root hair length and basal root growth angle for phosphorus acquisition. *Plant Physiol.* 167, 1430–1439. doi: 10.1104/pp.15.00145
- Miguel, M. A., Widrig, A., Vieira, R. F., Brown, K. M., and Lynch, J. P. (2013). Basal root whorl number: a modulator of phosphorus acquisition in common bean (*Phaseolus vulgaris*). *Ann. Bot.* 112, 973–982. doi: 10.1093/aob/mct164
- Multari, S., Stewart, D., and Russell, W. R. (2015). Potential of Fava bean as future protein supply to partially replace meat intake in the human diet. *Compr. Rev. Food Sci. Food Saf.* 14, 511–522. doi: 10.1111/1541-4337.12146
- Nicodemus, K. K., Callicott, J. H., Higier, R. G., Luna, A., Nixon, D. C., Lipska, B. K., et al. (2010). Evidence of statistical epistasis between DISC1, CIT and NDEL1 impacting risk for schizophrenia: biological validation with functional neuroimaging. *Hum. Genet.* 127, 441–452. doi: 10.1007/s00439-009-0782-y
- Okkan, U., and Serbes, Z. A. (2012). Rainfall–runoff modeling using least squares support vector machines. *Environmetrics* 23, 549–564. doi: 10.1007/s12011-010-8678-1
- Pan, X.-Y., and Shen, H.-B. (2009). Robust prediction of B-factor profile from sequence using two-stage SVR based on random forest feature selection. *Protein pept. lett.* 16, 1447–1454. doi: 10.2174/092986609789839250
- Pinheiro, H. A., DaMatta, F. M., Chaves, A. R., Loureiro, M. E., and Ducatti, C. (2005). Drought tolerance is associated with rooting depth and stomatal control of water use in clones of *Coffea canephora*. *Ann. Bot.* 96, 101–108. doi: 10.1093/aob/mci154
- Poorter, H., Jagodzinski, A. M., Ruiz-Peinado, R., Kuyah, S., Luo, Y., Oleksyn, J., et al. (2015). How does biomass distribution change with size and differ among species? An analysis for 1200 plant species from five continents. *New Phytol.* 208, 736–749. doi: 10.1111/nph.13571
- Preissel, S., Reckling, M., Schläpke, N., and Zander, P. (2015). Magnitude and farm-economic value of grain legume pre-crop benefits in Europe: a review. *Field Crops Res.* 175, 64–79. doi: 10.1016/j.fcr.2015.01.012
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. Vienna: R foundation for Statistical Computing.
- Rahaman, M. M., Chen, D., Gillani, Z., Klukas, C., and Chen, M. (2015). Advanced phenotyping and phenotype data analysis for the plant growth and development study. *Front. Plant Sci.* 6:00619. doi: 10.3389/fpls.2015.00619
- Rewald, B., Ephrath, J. E., and Rachmilevitch, S. (2011). A root is a root is a root? Water uptake rates of *Citrus* root orders. *Plant Cell Environ.* 34, 33–42. doi: 10.1111/j.1365-3040.2010.02223.x
- Rich, S. M., and Watt, M. (2013). Soil conditions and cereal root system architecture: review and considerations for linking Darwin and Weaver. *J. Exp. Bot.* 64, 1193–1208. doi: 10.1093/jxb/ert043
- Sacchet, M. D., Prasad, G., Folland-Ross, L. C., Thompson, P. M., and Gotlib, I. H. (2015). Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. *Front. Psychiatry* 6:00021. doi: 10.3389/fpsyt.2015.00021
- Sack, L., and Scoffoni, C. (2013). Leaf venation: structure, function, development, evolution, ecology and applications in the past, present and future. *New Phytol.* 198, 983–1000. doi: 10.1111/nph.12253

- Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit. Lett.* 33, 431–437. doi: 10.1016/j.patrec.2011.05.016
- Shang, X., and Chisholm, L. (2014). Classification of Australian native forest species using hyperspectral remote sensing and machine-learning classification algorithms. *JSTARS* 7, 2481–2489.
- Sommer, C., and Gerlich, D. W. (2013). Machine learning in cell biology—teaching computers to recognize phenotypes. *J. Cell Sci.* 126, 5529–5539. doi: 10.1242/jcs.123604
- Stachowicz, J. J., Kamel, S. J., Hughes, A. R., and Grosberg, R. K. (2013). Genetic relatedness influences plant biomass accumulation in eelgrass (*Zostera marina*). *Am. Nat.* 181, 715–724. doi: 10.1086/669969
- Statnikov, A., Hardin, D., and Aliferis, C. (2006). “Using SVM weight-based methods to identify causally relevant and non-causally relevant variables,” in *Proceedings of the NIPS 2006 Workshop on Causality and Feature Selection*, Whistler, BC.
- Steinberga, V., Alsina, I., Ansevica, A., Dubova, L., Liepina, L., and Šterne, D. (2008). The evaluation of effectiveness of *Rhizobium lupini* strains (In Latvian). *Agronomijas Vēstis* 10, 193–197.
- Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics* 9:307. doi: 10.1186/1471-2105-9-307
- Strobl, C., Malley, J., and Tutz, G. (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol. Methods* 14, 323–348. doi: 10.1037/a0016973
- Strobl, C., and Zeileis, A. (2008). *Danger: High power!—Exploring the Statistical Properties of a Test for Random Forest Variable Importance*. Munich: Ludwig-Maximilian University of Munich.
- Trachsel, S., Kaeppeler, S. M., Brown, K. M., and Lynch, J. P. (2011). Shovelomics: high throughput phenotyping of maize (*Zea mays* L.) root architecture in the field. *Plant Soil* 341, 75–87. doi: 10.1007/s11104-010-0623-8
- Tracy, S. R., Black, C. R., Roberts, J. A., Sturrock, C., Mairhofer, S., Craigon, J., et al. (2012). Quantifying the impact of soil compaction on root system architecture in tomato (*Solanum lycopersicum*) by X-ray micro-computed tomography. *Ann. Bot.* 110, 511–519. doi: 10.1093/aob/mcs031
- Tuberosa, R., Salvi, S., Sanguineti, M. C., Landi, P., Maccaferri, M., and Conti, S. (2002). Mapping QTLs regulating morpho-physiological traits and yield: case studies, shortcomings and perspectives in drought-stressed maize. *Ann. Bot.* 89, 941–963. doi: 10.1093/aob/mcf134
- Vadez, V., Rao, S., Kholova, J., Krishnamurthy, L., Kashiwagi, J., Ratnakumar, P., et al. (2008). Root research for drought tolerance in legumes: quo vadis. *J. Food Legumes* 21, 77–85.
- Vapnik, V. N., and Vapnik, V. (1998). *Statistical Learning Theory*, Vol. 1. New York, NY: Wiley.
- Vaz Patto, M. C., Amarowicz, R., Aryee, A. N. A., Boye, J. I., Chung, H.-J., Martín-Cabrejas, M. A., et al. (2014). Achievements and challenges in improving the nutritional quality of food legumes. *Crit. Rev. Plant Sci.* 34, 105–143. doi: 10.1080/07352689.2014.897907
- Verikas, A., Gelzinis, A., and Bacauskiene, M. (2011). Mining data with random forests: a survey and results of new tests. *Pattern Recognit.* 44, 330–349. doi: 10.1016/j.patcog.2010.08.011
- Wang, L., Huang, C., and Yang, J. Y. (2010). Predicting siRNA potency with random forests and support vector machines. *BMC Genomics* 11:S2. doi: 10.1186/1471-2164-11-S3-S2
- Watt, M., Moosavi, S., Cunningham, S. C., Kirkegaard, J. A., Rebetzke, G. J., and Richards, R. A. (2013). A rapid, controlled-environment seedling root screen for wheat correlates well with rooting depths at vegetative, but not reproductive, stages at two field sites. *Ann. Bot.* 112, 447–455. doi: 10.1093/aob/mct122
- Wojciechowski, T., Gooding, M. J., Ramsay, L., and Gregory, P. J. (2009). The effects of dwarfing genes on seedling root growth of wheat. *J. Exp. Bot.* 60, 2565–2573. doi: 10.1093/jxb/erp107
- Yang, J., Yao, D., Zhan, X., and Zhan, X. (2014). “Predicting disease risks using feature selection based on random forest and support vector machine,” in *Bioinformatics Research and Applications*, Vol. 8492, eds M. Basu, Y. Pan, and J. Wang (Berlin: Springer), 1–11.
- Zheng, B., Yoon, S. W., and Lam, S. S. (2014). Breast cancer diagnosis based on feature extraction using a hybrid of K-means and support vector machine algorithms. *Expert Syst. Appl.* 41, 1476–1482. doi: 10.1016/j.eswa.2013.08.044
- Zhong, D., Novais, J., Grift, T. E., Bohn, M., and Han, J. (2009). Maize root complexity analysis using a support vector machine method. *Comput. Electron. Agric.* 69, 46–50. doi: 10.1016/j.compag.2009.06.013
- Zhu, Z., Woodcock, C. E., Rogan, J., and Kellndorfer, J. (2012). Assessment of spectral, polarimetric, temporal, and spatial dimensions for urban and peri-urban land cover classification using Landsat and SAR data. *Remote Sensing Environ.* 117, 72–82. doi: 10.1016/j.rse.2011.07.020

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Zhao, Bodner and Rewald. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Identification of QTL and Qualitative Trait Loci for Agronomic Traits Using SNP Markers in the Adzuki Bean

Yuan Li^{1†}, Kai Yang^{1†}, Wei Yang², Liwei Chu¹, Chunhai Chen², Bo Zhao¹, Yisong Li¹, Jianbo Jian², Zhichao Yin³, Tianqi Wang¹ and Ping Wan^{1*}

¹ Key Laboratory of New Technology in Agricultural Application, College of Plant Science and Technology, Beijing University of Agriculture, Beijing, China, ² Beijing Genomics Institute-Shenzhen, Shenzhen, China, ³ College of Plant Science, Jilin University, Changchun, China

OPEN ACCESS

Edited by:

Maria Carlota Vaz Patto,
ITQB NOVA, Portugal

Reviewed by:

Liezhao Liu,
Southwest University, China
Pawel Krajewski,
Institute of Plant Genetics (PAN),
Poland

*Correspondence:

Ping Wan
pingwan3@163.com;
pingwan@bua.edu.cn

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 31 December 2016

Accepted: 04 May 2017

Published: 19 May 2017

Citation:

Li Y, Yang K, Yang W, Chu L, Chen C,
Zhao B, Li Y, Jian J, Yin Z, Wang T and
Wan P (2017) Identification of QTL and
Qualitative Trait Loci for Agronomic
Traits Using SNP Markers in the
Adzuki Bean. *Front. Plant Sci.* 8:840.
doi: 10.3389/fpls.2017.00840

The adzuki bean (*Vigna angularis*) is an important grain legume. Fine mapping of quantitative trait loci (QTL) and qualitative trait genes plays an important role in gene cloning, molecular-marker-assisted selection (MAS), and trait improvement. However, the genetic control of agronomic traits in the adzuki bean remains poorly understood. Single-nucleotide polymorphisms (SNPs) are invaluable in the construction of high-density genetic maps. We mapped 26 agronomic QTLs and five qualitative trait genes related to pigmentation using 1,571 polymorphic SNP markers from the adzuki bean genome via restriction-site-associated DNA sequencing of 150 members of an F₂ population derived from a cross between cultivated and wild adzuki beans. We mapped 11 QTLs for flowering time and pod maturity on chromosomes 4, 7, and 10. Six 100-seed weight (SD100WT) QTLs were detected. Two major flowering time QTLs were located on chromosome 4, firstly *VaFid4.1* (PEVs 71.3%), co-segregating with SNP marker s690-144110, and *VaFid4.2* (PEVs 67.6%) at a 0.974 cM genetic distance from the SNP marker s165-116310. Three QTLs for seed number per pod (*Snp3.1*, *Snp3.2*, and *Snp4.1*) were mapped on chromosomes 3 and 4. One QTL *VaSdt4.1* of seed thickness (SDT) and three QTLs for branch number on the main stem were detected on chromosome 4. QTLs for maximum leaf width (LFMW) and stem internode length were mapped to chromosomes 2 and 9, respectively. Trait genes controlling the color of the seed coat, pod, stem and flower were mapped to chromosomes 3 and 1. Three candidate genes, *VaAGL*, *VaPhyE*, and *VaAP2*, were identified for flowering time and pod maturity. *VaAGL* encodes an agamous-like MADS-box protein of 379 amino acids. *VaPhyE* encodes a phytochrome E protein of 1,121 amino acids. Four phytochrome genes (*VaPhyA1*, *VaPhyA2*, *VaPhyB*, and *VaPhyE*) were identified in the adzuki bean genome. We found candidate genes *VaAP2/ERF.81* and *VaAP2/ERF.82* of SD100WT, *VaAP2-s4* of SDT, and *VaAP2/ERF.86* of LFMW. A candidate gene *VaUGT* related to black seed coat color was identified. These mapped QTL and qualitative trait genes provide information helpful for future adzuki bean candidate gene cloning and MAS breeding to improve cultivars with desirable growth periods, yields, and seed coat color types.

Keywords: adzuki bean (*Vigna angularis*), agronomic trait, QTL, qualitative trait, SNP marker, candidate gene

INTRODUCTION

The adzuki bean (*Vigna angularis*) is an important diploid pulse crop ($2n = 2x = 22$) that is rich in easily digestible protein with extremely low fat content (Lin, 2002). It was domesticated about 12,000 years ago in China (Liu et al., 2013), and is cultivated today in over 30 countries of the world, principally those of eastern and northern Asia (Tomooka et al., 2002; Kramer et al., 2012). China is the largest producer of adzuki beans in the world, with an area of approximately 25,000 ha cultivated annually (Cheng and Tian, 2009). The adzuki bean is a rich source of phenolic compounds, flavonoids, vitamin A, vitamin B, iron, zinc, and folate (Amarowicz et al., 2008; Yao et al., 2012).

Gene and QTL mapping is very important for gene cloning, MAS breeding, and trait improvement; however, only a few studies have focused on mapping the QTL and the qualitative trait genes in the adzuki bean. The QTLs of 21 domestication-related traits were first mapped to different linkage groups by 21 polymorphic SSR markers using the same BC₁F₁, F₂, and F_{2:3} populations to construct a molecular linkage map (Han et al., 2005). Most traits mapped to particular regions of linkage groups (LGs) 1, 2, 4, 7, and 9. Pod size, germination efficacy, seed size, and lower stem length mapped to LGs 1 and 2. The QTLs of LGs 7 and 9 were associated with upper-stem length, maximum leaf size, and pod and seed sizes (Isemura et al., 2007).

Kaga et al. (2008) used 316 SSR primer pairs from the adzuki bean (Wang et al., 2004), 170 SSR primer pairs from the common bean, 45 cowpea SSR primer pairs, and AFLPs to screen for polymorphisms in the two parents. In total, 176 adzuki bean SSRs and 5 common bean SSR primer pairs exhibited clear polymorphisms. F₂ mapping population consisted of 188 plants derived from crosses between the Japanese wild bean (*V. angularis* var. *nipponensis*) and the cultivated adzuki bean (*V. angularis* var. *angularis*). The AFLP approach was developed to fill a large gap (~40 cM) in the center of LG9. In total, 233 markers (191 SSRs, 2 STSs, 1 CAPS, 2 SCARs, and 36 AFLPs) and three morphological traits were mapped to 10 linkage groups (one less than the 11 haploid chromosomes of the adzuki bean). One linkage group, termed “LG4+6,” contained the LG4 and LG6 markers of previous maps. In total, 162 QTLs influencing 46 domestication-related traits were identified. The QTLs affecting seed dormancy; seedling stem length; red seed-coat color; and the organ sizes of seeds, pods, and leaves, were mapped to LG1. The QTLs for pod dehiscence, length, size, and color lay on LG7. The QTLs for organ size, growth habit, and yield-related traits (total seed weight, total pod and seed numbers, 100-seed weight, and total seed weight), maximum leaflet length, primary leaf width, and pod width and length lay in two distinct regions of LG9 (Kaga et al., 2008).

Azuki Dwarf1 (AD1), a single genetically unstable dwarf locus, co-segregates with SSRs on LG4 (Aoyama et al., 2011). A strong QTL for seed coat color, designated *OLB1*, explains 54.43 and 56% of the total variances in the L* (lightness), a* (redness), and b* (yellowness) values. In addition, a minor QTL, designated *OLB2*, explains 6% of the total variance in redness. *OLB1* and *OLB2* are located in LG1. Furthermore, two traits controlled by a single Mendelian gene: *IVY* (ivory/yellow) and *POB* (pale

olive/buff) (seed coat colors) are located in LG8 and LG10, respectively (Horiuchi et al., 2015).

We previously published a draft version of the adzuki bean genome (Yang et al., 2015), which will facilitate the identification of agronomic trait genes and accelerate the improvement of adzuki bean.

However, to the best of our knowledge, no QTL analysis using high-density segregated SNPs has been performed in adzuki bean. In this study, we initially collected phenotypic data and then defined genotypic data using SNP markers via restriction-site-associated DNA (RAD) sequencing. The QTLs of important agronomic traits of the adzuki bean were mapped using these polymorphic SNP markers. Our results elucidate how genetic features control the agronomic traits of the adzuki bean, and the major QTLs and genes that we have identified will expedite MAS breeding and the improvement of these traits in adzuki bean.

MATERIALS AND METHODS

The Mapping Population

The F₂ mapping population initially comprised 250 individuals derived by crosses between an adzuki bean cultivar (Ass001) and a wild adzuki bean (accession # CWA108) collected in China. The F₂ population, and 10 plants of each parent, were grown in the Experimental Farm of Beijing University of Agriculture (BUA) from June to October, 2013. A single seed was planted with 60-cm row spacing and 30-cm plant spacing. Each F_{2:3} line had two rows that were 3 m long and 45 cm wide; 35 seeds were evenly planted in each row in the field in June 2014 at the BUA Experimental Farm. From the center of the rows, 10 representational plants were selected to evaluate traits, and 10 plants from each parent were grown together with the F_{2:3} line. The RIL (Recombinant Inbred lines) of F₃ were obtained from single seed descendent of the F₂ individuals; their parents were grown in the same manner as F₂ at the BUA Experimental Farm in 2014 and 2015. F_{3:4} lines and parents were planted and evaluated for traits in the same manner as F_{2:3} lines at the BUA Experimental Farm in 2015.

We selected 150 typical F₂ individuals to extract DNA for RAD sequencing.

Trait Measurements

In total, 28 traits, including 24 quantitative, and 4 qualitative traits, were evaluated in F₂, F₃, and F₄ generations from 2013 to 2015 (Table 1; Table S1). Morphology was investigated according to a published standard (Tomooka et al., 2002; Cheng et al., 2012). The traits included the color of the seed-coat, pod, stem and flower; the first-to-tenth internodal length; the maximal leaf length and width; the maximum leaflet area; and the growth habit at 50% of flowering; we investigated plant height, stem diameter, the number of branches, flowering time, and pod maturing time; and 100-seed weight, seed size, and pod size were measured after harvesting, respectively. The maximal leaf lengths and widths, and leaflet areas, were estimated with the aid of a YMJ-C leaf area meter (Zhejiang Top Instrument Co., Ltd., China). The lengths, widths, and thicknesses of 10 seeds were measured using digital calipers.

TABLE 1 | Traits examined and the evaluation method.

Trait	Trait abbreviation	Evaluation method
Plant height (cm)	PHT	Plant height from cotyledonary node to tip of mature plant
Stem thickness (cm)	STT	Stem diameter below the primary leaf
Branch position of first branch	BRP	Position of 1st branch on main stem from 1st trifoliate leaf node to 10th trifoliate leaf node
Branch number on main stem	BRN	Number of branches on main stem from 1st trifoliate leaf node to 10th trifoliate leaf node
Stem internode length (first to tenth) (cm)	ST1I-ST10I	Length from primary leaf node to each node
Pod length (cm)	PDL	Length of straight pod
Pod width (cm)	PDW	Width of pod at widest part
Pod number per plant	PDTN	Pod number per plant (10 plants)
Seed number per pod	SDNPPD	Seed number per pod (10 pods)
Total seed number.	SDTN	Seed number per plant (10 plants)
Total seed weight	SDTWT	Seed weight per plant (10 plants)
100 seed weight (g)	SD100WT	Weight of 100 seeds
Seed length (mm)	SDL	Maximum distance from top to bottom of the seed
Seed width (mm)	SDW	Maximum distance from hilum to its opposite side
Seed thickness (mm)	SDT	Maximum distance between the two sides of the hilum
Pod color	PDC	Black, straw
Seed coat color	SDC	Black, brown, red
Growth habit	GH	Stem twining beyond the main stem 10th internode (erect, semi-erect, twining)
Stem color	STC	Purple, green
Days to first flower (days)	FLD	Number of days from sowing to first flowering
Days to 50% flower (days)	FLD50	Number of days from sowing to 50% flowering
Days to maturity of 25% pods (days)	PDDM25	Number of days from sowing to 25% mature pods
Days to maturity of 50% pods (days)	PDDM50	Number of days from sowing to 50% mature pods
Days to maturity of 75% pods (days)	PDDM75	Number of days from sowing to 75% mature pods
Days to maturity of 100% pods (days)	PDDM100	Number of days from sowing to 100% mature pods
Maximum leaf area (cm ²)	LFMA	Leaflet area of the largest terminal leaflet on a leaf between 1st trifoliate leaf node and 10th trifoliate leaf node
Maximum leaf length (cm)	LFML	Length of largest terminal leaflet on a leaf between 1st trifoliate leaf node and 10th trifoliate leaf node
Maximum leaf width (cm)	LFMW	Width of largest terminal leaflet on a leaf between 1st trifoliate leaf node and 10th trifoliate leaf node

RAD-Sequencing and SNP Detection

Genomic DNA of 150 F₂ individual plants and 10 plants of their parents were selected from which to extract genomic DNA from young leaves using the CTAB protocol. DNA was digested with *EcoRI* using the method of Baird et al. (2008) with minor modifications for RAD sequencing. SNP detection and genetic map construction followed the method of Yang et al. (2015).

Data Analysis and QTL Mapping

We calculated the mean, the standard error of the mean, and broad-sense heritability (the H_B^2 value) of investigated 24 quantitative traits, including plant height, organ size, yield, flowering time and maturing period, and plant architecture (Table 1) for parents, and the F₂, F₃, F₄ populations. The heritability was computed using a method of regression for progeny values on parental values. The calculation of H_B^2 employed the formula: $H_B^2 = b_{F_2:F_2:3} = COV_{F_2:F_2:3} / V_{F_2}$.

Using RAD tag technology, SNPs were identified among 150 F₂ individuals derived from a cross of cultivar Ass001 (P1) with the wild adzuki bean accession no. CWA108 (P2). After genotyping, a linkage map was created using *JoinMap 4.0* software (Van Ooijen, 2006) running F₂ population-type codes.

Markers exhibiting distorted segregation ($p < 0.01$, Chi-squared test) were excluded (Grattapaglia and Sederoff, 1994). The remaining 1,571 markers were used to construct a genetic map. Eleven linkage groups (for adzuki bean, $2n = 22$ chromosomes) were formed with the logarithm-of-the-odds (LOD) score set to 6.0 (Yang et al., 2015) and ordered using a regression mapping algorithm. Recombination frequencies were translated to genetic distances using Kosambi's mapping function (Kosambi, 1994). Qualitative trait genes were mapped using *JoinMap 4.1*.

Five Mendelian phenotype markers (stem color, flower color, pod color, and two seed coat colors) were detected as described in the Methods section. These phenotype markers were used as molecular markers for genotyping and linkage grouping.

Further QTL analysis was performed using MapQTL 6.0 software, which is widely used in the analysis of QTLs (Van Ooijen, 2009). First, the PERMUTATION test was used to obtain genome-wide LOD thresholds ($p < 0.05$), and each trait was subjected to 10,000 permutations to derive the empirical LOD threshold (Churchill and Doerge, 1994). Next, the regression approach of the interval mapping model was introduced to obtain LOD values for all the significant markers, and these were associated with candidate traits. All mapping markers for

which the LOD value was equal to or greater than the LOD threshold value were retained. Finally, these significant markers were used as cofactors in the multiple QTL method (MQM) (Jansen, 1993), such that the identified markers were adjacent to the significant QTLs in each group. All mapping information including chromosomal location, magnitude, direction of the additive effect, and the proportion of the phenotypic variation explained (PVE) in each detected QTL was obtained from the MQM outputs. Markers with LODs greater than the LOD threshold were identified and were regarded as the optimal final markers.

Candidate Gene Identification and Phylogenetic Analysis

Based on candidate gene sequences of *E3*, which encodes phyA-type photoreceptors involved in the control of flowering and maturity in soybean (Liu et al., 2008; Watanabe et al., 2009), transcription factor agamous-like MADS-box and *AP2* were detected in the mapping regions for flowering time, maturity, and seed coat color. We downloaded 427 *AP2* gene sequences, 123 MADS gene sequences, 95 *PHY* gene sequences, and 150 *UGT* gene (UDP flavonoid glycosyl transferase) sequences of *Arabidopsis thaliana* from NCBI. All of these protein sequences were used as seed sequences during gene copy number analysis. These four gene types were searched for and identified from the adzuki bean genome (PRJNA261643), and other sequenced legume genomes like soybean (phytozomev10), chickpea (<http://gigadb.org/dataset/100076>), *Medicago truncatula* (phytozomev10), pigeonpea (<http://gigadb.org/dataset/100028>), common bean (phytozomev10), mung bean (<ftp.ncbi.nih.gov>) and *Arachis duranensis* (<http://www.peanutbase.org/>), and from the genomes of *Arabidopsis* (TAIR9, Phytozome v10.0), *Brassica rapa* (Phytozome v10.0), rice (<http://rice.plantbiology.msu.edu/>), maize (phytozomev10). Previously published related gene sequences from *Arabidopsis* genomes were collected and used as query sequences. These query sequences were then used to align each sequenced legume genome sequence using TBLASTN v2.2.23 with a threshold *E*-value less than 10^{-10} . Because we obtained so many alignment results within the nearby genomic area, we extracted high quality alignments (query_align_ratio, the ratio of alignment length of query sequences size $\geq 70\%$ and identity $\geq 40\%$). Functional intact genes were confirmed via collection of blast-hits using the above method. Each of the blast-hit sequences was then extended in both 3' and 5' directions along the genome sequences to predict gene structure using Genewise. The resulting sequences were further confirmed by phylogenetic structure analysis. Finally, the coding sequences with proper ATG or the stop codon were extracted, but not those with interrupting stop codons or frame shifts.

RESULTS

Phenotypic Variation and Genetic Analysis

The phenotypes of the cultivated and wild plants differed significantly. The female parent Ass001 (a cultivar) had a red seed coat, a straw-colored pod, a green stem, an erect plant,

and a large seed. The wild parent had a black seed coat, a black pod, a purple stem, a twining plant, and a small seed. The different individuals in the same generation exhibited wide segregation for agronomic traits or variation (Table S1, Table 2, Figure 1).

We calculated the trait means, the standard error of the mean, and the heritability of the parental F_2 , F_3 , and F_4 populations (Table 2). The leaf size, seed and pod sizes, seed number per pod (SDNPPD), seed total number (SDTN), and seed total weight (SDTWT) of the cultivated parent were greater than those of the wild parent. The mean values of the wild parent were greater than those of the cultivated parent for plant height, branch number on the main stem, stem internode length (from the first to the tenth internode [ST1I-ST10I]), and flowering and maturation times. The individuals in F_2 showed wide segregation, and different lines in F_3 or F_4 exhibited similar phenotypic variation (Table 2). The means of various parameters of the F_2 , F_3 , and F_4 populations generally lay between those of the cultivated and wild parents, except for plant height, ST1I-ST10I, the total number of pods per plant, days of $\geq 50\%$ flowering (F_2), and plant height, total number of pods per plant, and SDTN (F_3). Many traits were highly heritable ($>70\%$).

Seed coat color, pod color, and stem color are all qualitative traits. The segregation ratios of the pod color and stem color were as expected (3:1) based on the Chi-squared test. The seed coat colors of the F_2 population included black, light brown, and red. The Chi-squared test showed that the segregation ratio of black: light brown: red seed coats was consistent with a 12:3:1 ratio; these seed coat colors were controlled by two genes. Black to light brown showed dominant epistasis (12:3, $\chi^2 = 3.820 < \chi^2_{0.05} = 3.84$), light brown was dominant to red (3:1, $\chi^2 = 0.701 < \chi^2_{0.05} = 3.84$).

QTL Detection and Analysis

We previously constructed a high-density SNP genetic map that we used only in genome assembly to anchor the scaffolds of the adzuki bean genome to the chromosomes (Yang et al., 2015); however, no morphological characteristics or QTL were involved in that SNP genetic map. The SNP genetic map was composed of 1,571 SNPs covering 11 linkage groups, spanning 1,031.17 cM, with an average of 4.33 mapped SNPs per scaffold at a mean marker distance of 0.67 cM (Table S2, Table 3). We used 1,571 polymorphic SNP markers to map QTL and qualitative trait genes in this study. In total, we identified 26 QTLs for flowering time, growth period, agronomic traits, and yield traits (Figures 2–4 and Table 4).

Flowering Time and Growth Period

We found 11 QTLs affecting flowering time and pod maturity on chromosomes 4, 7, and 10 (Figure 2). Most flowering and maturation time genes mapped to chromosome 4, except for the two minor QTLs *Fld4.5* and *Fld4.6*. The “days to first flowering” trait was controlled by only a few genes. Two major QTLs (*VaFld4.1* and *VaFld4.2*; PEVs 71.3% and 67.6, respectively) were identified in the 9.65 cM region of chromosome 4 of the F_2 population affecting flowering and maturation times. Two “first flowering” QTLs with smaller effects were found on

TABLE 2 | The mean, standard error of mean, and heritability values for parents, the F₂, F₃, F₄ populations of the cross between cultivated and wild adzuki bean.

	F ₂ population						
	Ass001		CWA108		F ₂		
	Mean	SEM	Mean	SEM	Mean	SEM	CV
PHT	50.60	1.81	67.80	5.93	88.70	3.20	42.66
STT	0.57	0.02	0.55	0.01	0.35	0.01	39.35
BRP	1.44	0.24	1.40	0.25	1.99	0.11	64.59
BRN	5.60	0.58	10.90	0.38	6.33	0.19	34.71
ST1I–ST10I	12.05	0.48	16.70	1.55	18.43	0.59	38.01
PDL	7.70	0.17	3.57	0.19	5.50	0.08	17.94
PDW	0.65	0.01	0.29	0.01	0.48	0.00	11.31
PDTN	25.30	3.56	20.56	1.89	26.38	2.08	93.38
SDNPPD	4.84	0.21	2.67	0.44	4.18	0.10	28.58
SDTN	123.20	18.67	54.22	8.84	118.18	10.19	102.04
SDTWT	17.11	2.54	0.90	0.13	8.34	0.75	106.13
SD100WT	13.95	0.30	1.69	0.12	6.77	0.14	24.00
SDL	0.70	0.01	0.36	0.01	0.52	0.00	10.27
SDW	0.53	0.01	0.27	0.01	0.44	0.00	8.14
SDT	0.56	0.00	0.35	0.00	0.41	0.00	7.81
FLD	57.20	1.53	81.30	0.30	77.96	1.31	19.93
FLD50	67.40	1.42	87.70	0.26	89.10	1.28	16.99
PDDM25	91.70	0.60	112.80	0.68	103.32	1.24	13.94
PDDM50	97.20	0.94	120.80	1.01	111.33	1.25	12.92
PDDM75	105.20	1.04	122.50	0.98	117.85	1.19	11.14
PDDM100	114.10	0.89	138.20	2.15	126.13	1.16	10.39
LFMA	41.47	1.54	5.72	0.50	23.45	0.67	33.49
LFML	10.02	0.22	4.28	0.18	7.61	0.13	19.39
LFMW	6.87	0.14	2.76	0.14	5.26	0.09	19.56
PDC	Straw		Black				
SDC	Red		Black				
GH	Erect		Twining				
STC	Green		Purple				

	F ₃ population							Heritability F ₃ on F ₂
	Ass001		CWA108		F ₃			
	Mean	SEM	Mean	SEM	Mean	SEM	CV	
PHT	42.20	2.18	64.20	8.51	73.96	3.89	54.41	24.98
STT	0.55	0.03	0.56	0.02	0.38	0.05	27.47	34.83
BRP	2.80	0.44	2.00	0.45	2.15	0.14	67.57	0.00
BRN	4.00	0.49	11.40	0.51	5.30	0.28	54.45	36.02
ST1I–ST10I	15.80	0.53	27.40	1.44	17.83	0.53	30.87	32.04
PDL	8.14	0.06	4.46	0.38	5.80	0.09	15.16	16.77
PDW	0.62	0.01	0.24	0.01	0.46	0.00	9.39	1.64
PDTN	13.90	1.93	19.60	2.38	19.99	1.92	99.35	6.80
SDNPPD	5.19	0.13	5.00	0.63	5.06	0.14	28.21	6.95
SDTN	72.90	7.51	95.00	12.03	111.20	11.86	110.30	12.72
SDTWT	12.70	1.07	1.33	0.17	7.25	0.99	140.50	7.79
SD100WT	17.51	0.85	1.36	0.04	5.78	0.18	32.95	45.70
SDL	0.76	0.01	0.41	0.01	0.64	0.05	74.56	50.81
SDW	0.63	0.01	0.28	0.01	0.05	0.02	453.79	35.44
SDT	0.58	0.01	0.26	0.01	0.39	0.00	12.26	32.48
FLD	54.80	0.49	92.00	0.71	66.49	0.94	14.55	34.46
FLD50	62.20	0.49	99.00	0.95	77.75	0.85	11.35	21.96

(Continued)

TABLE 2 | Continued

	F ₃ population							Heritability F ₃ on F ₂
	Ass001		CWA108		F ₃			
	Mean	SEM	Mean	SEM	Mean	SEM	CV	
PDDM25	76.00	0.54	128.00	0.84	82.60	0.97	9.79	18.32
PDDM50	90.00	0.56			91.80	0.78	8.31	13.62
PDDM75	100.00	0.65	135.20	1.24	99.86	1.00	8.42	13.39
PDDM100	106.00	0.79	140.00	2.49	105.78	0.95	9.32	9.75
LFMA	42.72	1.14	6.20	0.80	23.71	0.73	31.96	89.20
LFML	10.35	0.17	4.48	0.30	7.64	0.15	19.69	97.08
LFMW	7.06	0.13	2.93	0.22	5.23	0.10	20.02	98.39
PDC	Straw		Black					
SDC	Red		Black					
GH	Erect		Twining					
STC	Green		Purple					

	F ₄ population							Heritability F ₄ on F ₃
	Ass001		CWA108		F ₄			
	Mean	SEM	Mean	SEM	Mean	SEM	CV	
PHT	43.75	1.42	50.20	7.14	36.40	1.40	40.05	3.68
STT	0.63	0.04	0.45	0.01	0.42	0.01	28.04	31.68
BRP	2.00	0.32	2.00	0.32	2.46	0.14	60.41	0.00
BRN	3.80	0.73	5.20	0.58	3.20	0.22	70.48	31.76
ST1I-ST10I	15.80	1.28	20.00	1.30	16.39	0.55	34.71	29.51
PDL	10.43	0.20	3.92	0.35	6.86	0.16	23.55	91.78
PDW	0.80	0.01	0.24	0.01	0.50	0.01	25.65	0.01
PDTN	17.20	1.98	16.00	2.19	21.10	1.77	86.67	35.09
SDNPPD	7.42	0.37	4.60	0.40	6.75	0.19	29.38	48.58
SDTN	90.20	9.85	88.00	6.00	104.92	10.15	99.57	33.09
SDTWT	18.82	2.56	1.63	0.22	8.14	1.00	126.76	57.17
SD100WT	20.91	0.51	2.25	0.17	6.89	0.23	34.65	73.59
SDL	0.83	0.01	0.40	0.01	0.52	0.01	12.88	52.32
SDW	0.59	0.02	0.30	0.01	0.39	0.00	12.07	53.83
SDT	0.65	0.01	0.36	0.00	0.41	0.01	15.67	55.15
FLD	50.00	1.52	95.00	0.32	54.38	0.91	17.74	43.72
FLD50	59.00	1.36	104.00	0.35	65.99	1.12	17.84	47.67
PDDM25	105.00	0.89	130.00	0.95	103.12	0.65	6.63	31.53
PDDM50								
PDDM75								
PDDM100	115.00	1.05	149.00	1.14	113.48	0.64	5.96	0.47
LFMA			6.13	0.56	24.35	0.73	31.63	94.03
LFML			4.13	0.15	7.85	0.14	18.22	93.31
LFMW			2.67	0.15	5.34	0.09	18.52	91.38
PDC	Straw		Black					
SDC	Red		Black					
GH	Erect		Twining					
STC	Green		Purple					

Trait abbreviations are shown in **Table 1**. Populations of trait value was listed. SEM, standard errors of mean values. CV, coefficient of variation.

chromosome 4. The SNP marker s690-144110 co-segregated with *VaFld4.1*, and the genetic distance between *VaFld4.2* and the SNP marker s165-116310 was 0.974 cM. The *VaFld4.2* flowering time QTL was present in the same chromosome 4 locus in the F₄

population, and co-segregated with the s165-116310 SNP marker. Two FLD50 ("time to 50% flowering") QTLs were located in the scaffold regions of *VaFld4.1* (PEV 73.1%) and *VaFld4.2* (PEV 69.2%), and two other minor flowering time QTLs were located

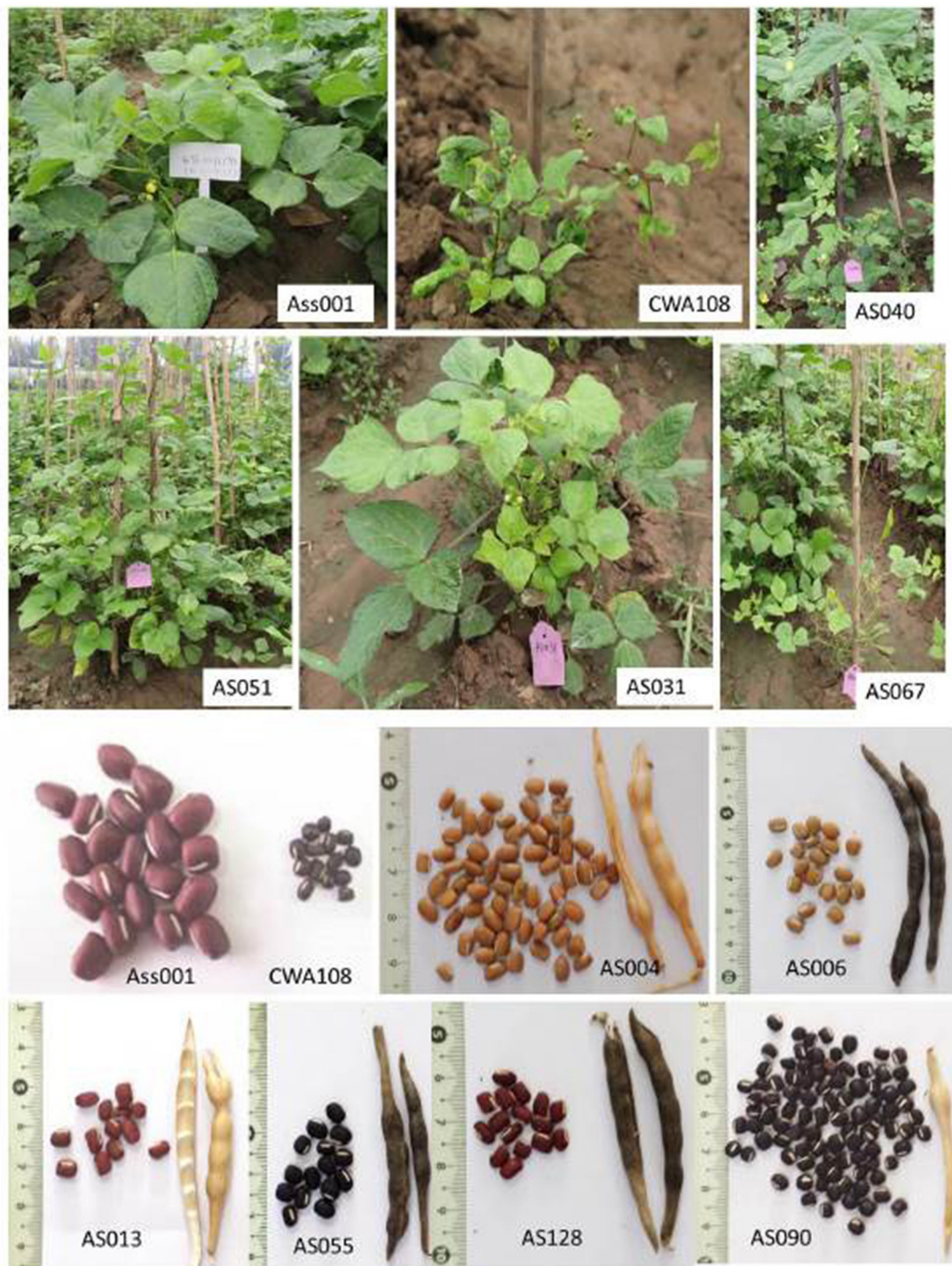


FIGURE 1 | Phenotype of parents (Ass001 and CWA108) and several individuals in the F_2 mapping population Ass001: female parent; CWA108: male parent; AS: individuals of F_2 .

on chromosome 10, with PEVs of 3.9%. One minor QTL was mapped on chromosome 7. PDDM25 (“25% pod maturation”, PEV 55.4%), PDDM50 (“50% pod maturation”, PEV 64.9%), and PDDM75 (“75% pod maturation”, PEV 71.3%) mapped to the same locus *VaFld4.1*, and co-segregated with the s690-144110

SNP marker (Table 4, Figure 2). The remaining two major PDDM75 QTLs, *VaGp4.1* and *VaGp4.2*, mapped to the 2.76 cM region of chromosome 4 (PEVs 47 and 40.9%). The *VaGp4.1* and *VaGp4.2* QTLs for FLD and the *VaGp4.1* QTL for FLD50 were identified in the F_3 generation.

TABLE 3 | Summary of the high-density SNP genetic map of adzuki beans.

Linkage group ID	Total SNP number	Total distance (cM)	Anchored scaffold length (bp)	Mbp/cM
Chr 1	64	128.24	42,245,285	0.33
Chr 2	135	108.42	35,639,406	0.33
Chr 3	172	108.30	41,220,908	0.38
Chr 4	108	104.42	34,681,404	0.33
Chr 5	194	101.70	30,704,717	0.30
Chr 6	135	98.59	19,363,203	0.20
Chr 7	164	95.13	39,418,547	0.41
Chr 8	97	85.72	30,803,815	0.36
Chr 9	111	84.48	31,756,850	0.38
Chr 10	255	70.45	40,442,837	0.57
Chr 11	136	45.72	26,600,026	0.58
Total	1,571	1031.17	372,876,998	–

Morphology and Yield Traits

In total, six SD100WT QTLs were detected, four on chromosome 1 and two on chromosome 11 (**Figure 3**). *VaSd100wt1.1* and *VaSd100wt1.2* were mapped to chromosome 1. *VaSd100wt1.1* located between SNP markers s268-1701932 and s268-1618129 with 0.0420 cM genetic distance on chromosome 11 (**Table 4**). Three seed-number-per-pod (SDNPPD) QTLs were identified on chromosome 3 (*Snp3.1* PEVs 17.3%) and chromosome 4 (*Snp4.1*, PEV 11.8%). *Snp3.1* was located at a distance of 0.012 cM between SNP markers s624-5207 and s624-5245, and *Snp3.2* lay on the same mapping region of *Snp3.1*. *Snp4.1* co-segregated with s624-5245 and s80-1776704, respectively. Three QTLs for branch number on the main stem mapped to chromosome 4 (**Figure 3**, **Table 4**). *VaBrn4.1* was found on chromosome 4 between SNP markers s690-144110 and s856-165880, and the genetic distance of *VaBrn4.2* was 0.974 cM with a s165-116310 marker (**Figure 4**). A QTL *VaLfmw2.1* for maximum leaf width (LFMW) co-segregated with a s536-414880 marker on chromosome 2 (PEV 21.2%). A QTL *VaST1-1019* of stem internode length from the first to the tenth node (ST11-ST101) was found on chromosome 9 (PEV 17.3%) and co-segregated with the SNP marker s249-3204596 (**Table 4**, **Figure 4**). QTL *VaSdt4.1* of SDT (seed thickness) was mapped on chromosome 4 between SNP markers s58-1489334 and s58-345785, and co-segregated with an S58-1489334 marker (**Figure 4**, **Table 4**).

Qualitative Trait Gene Mapping

Pigmentation-related genes controlling seed coat color (SDC), pod color (PDC), stem color (STC), and flower color (FLC) were mapped to chromosome 3. *VaFcY* controlling yellow flower color mapped to the top of chromosome 3, followed by the black seed coat color gene *VaScB*, and the green stem color gene *VaStcG* (**Figure 5**). The genetic distances between *VaStcG*, *VaScB*, *VaFcY*, and the SNP marker s342-127390 were 8.82, 12.95, and 41.77 cM, respectively. The black pod *VaPcB* gene was located between SNP markers s225-928306 and s101-825050, the genetic distances of which were 18.16 and 19 cM, respectively. The red seed coat color *VaScR* gene mapped on top of chromosome 1.

Identification of Candidate Genes

Flowering Time and Pod Maturity

Two flowering time candidate genes, *E3*-like phytochrome and transcription factor agamous-like MADS-box loci, and an *AP2* locus were detected in relative mapping regions controlling flowering time, maturity, and seed size.

VaPhyE (phytochrome E) encodes a protein of 1,121 amino acids. It is located in the interval between 3,182,263 and 3,186,619 bp on chromosome 4. In total, four phytochrome genes—two *VaPhyA*, a *VaPhyB*, and a *VaPhyE* gene—were found in the adzuki bean genome (**Table S3**). Based on the protein sequence encoded by the *VaPhyE* gene, clustering results showed that adzuki bean *VaPhyE* and mung bean *VrPhyE* have the closest genetic relationship, followed by the common bean *PvPhyE* (**Figure 6**).

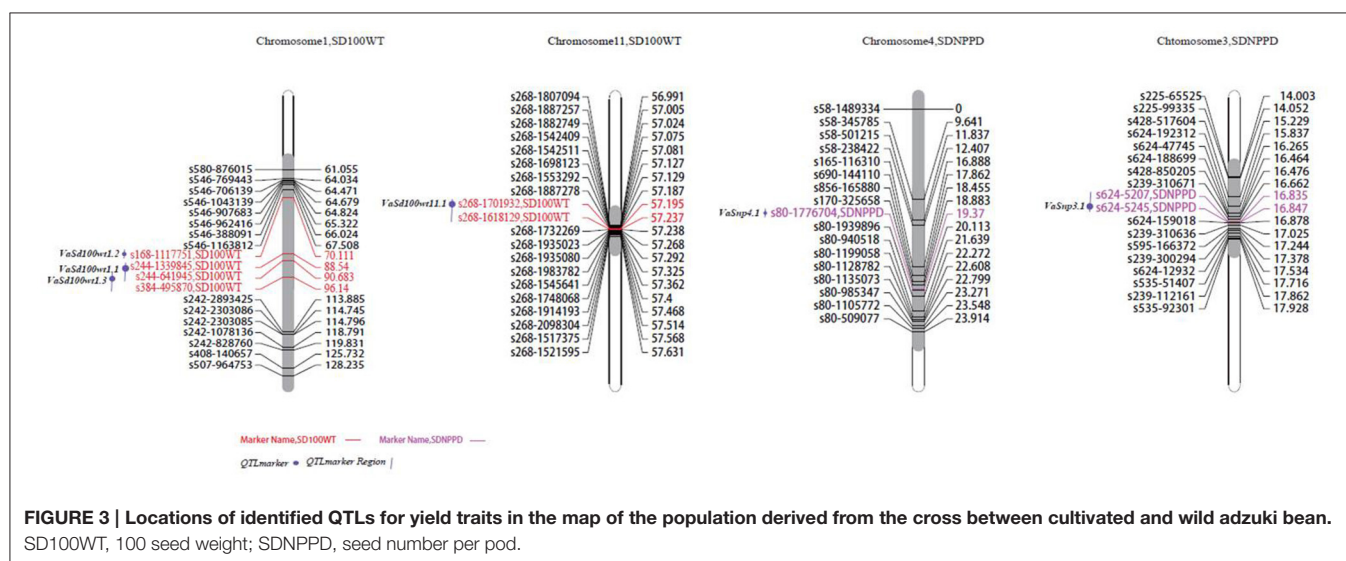
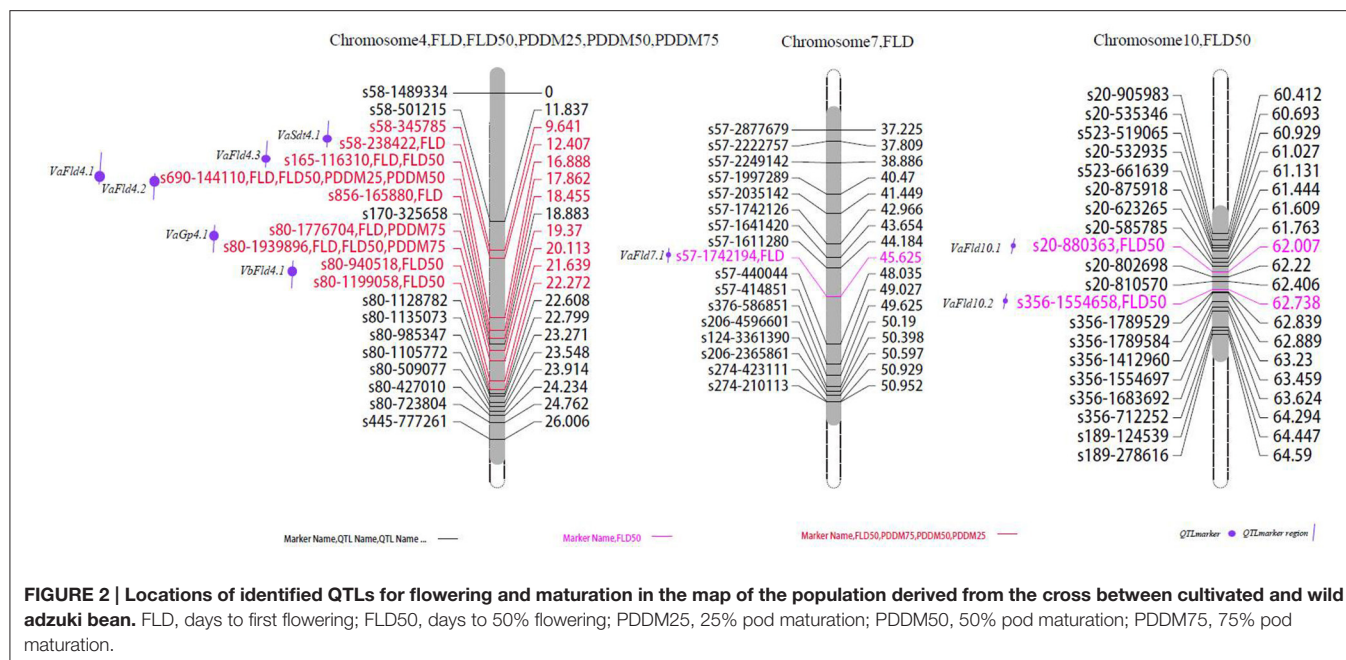
The agamous-like MADS-box is involved in flower development and maturity (de Folter et al., 2005). Two major QTLs controlling flowering time and pod maturity traits (FLD, FLD50, PDDM50, and PDDM100), *VaFld4.1* and *VaFld4.3*, were mapped in the interval from 3,102,255 to 3,616,262 bp on chromosome 4 between SNP markers s690-144110 and s856-165880; s856-165880 was the closest. An agamous-like MADS-box candidate gene, *VaAGL*, encoding a protein of 379 amino acids, was detected in this region. In total, 29 *Agamous*-like MADS-box genes were found in the adzuki bean genome (**Table S4**).

Seed and Leaf Size

Two *VaAP2/ERF.81* and *VaAP2/ERF.82* candidate genes encoding 278 and 225 amino acids were identified in the SD100WT QTL *VaSd100wt1.2* region of 28,861,763–28,863,078 bp and 29,072,132–29,072,809 bp on chromosome 1, and the closest SNP marker was s168-1117751 (**Figure 3**). The candidate gene *VaAP2-s4* was found in QTL *VaSdt4.1* of SDT (seed width) from 345,785 to 1,489,334 bp on chromosome 4 between SNP markers s58-345785 and s58-1489334 (**Figure 4**). *VaAP2-s4* encodes a protein of 208 amino acids. Another candidate gene, *VaAP2/ERF.86*, encoding 638 amino acids, was identified in the mapping region of QTL *VaLfmw2.1* (LFMW, maximum leaf width) from 20,746,593 to 20,747,231 bp on chromosome 2, between SNP markers s536-414880 and s536-414880 (**Figure 4**). In total, 26 *AP2/ERF* genes were identified in the adzuki bean genome (**Table S5**). *AP2/RAV* genes were absent in adzuki beans, common beans, mung beans, chickpeas, pigeonpeas, and *Medicago truncatula*, but did exist in soybeans.

Seed Coat Color Genes

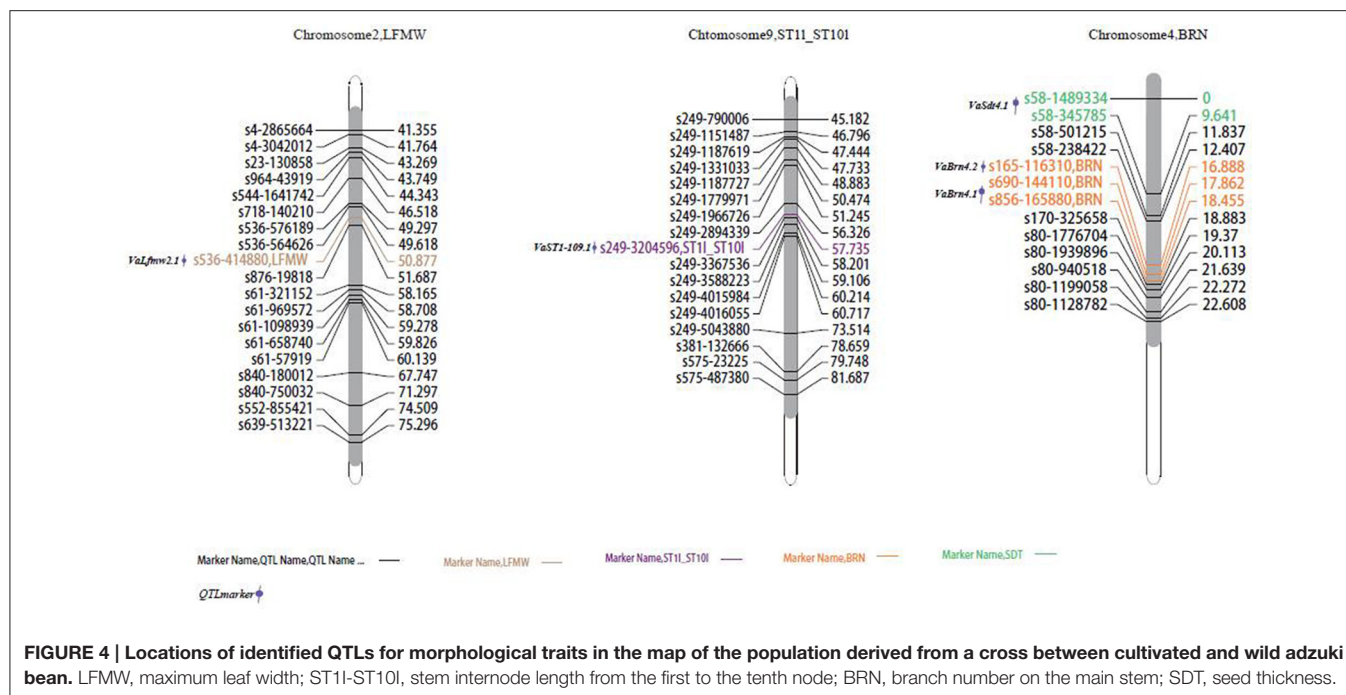
The *VaScB* gene controlling the black seed coat trait (SDC1) was mapped onto the interval from 131,943 to 133,424 bp on chromosome 3 between SNP marker scaffold326-733037 (28.83 cM) and the top of the chromosome 3. A candidate gene, *VaUGT*, was found within this interval. *VaUGT* encodes a protein of UDP flavonoid glycosyl transferase (UGT), with 494 amino acids, associated with flavonoid metabolic pathways. In total, 22 *UGT*-like genes were found in the adzuki bean genome (**Table S6**, **Figure S1**).



DISCUSSION

Gene mapping and QTL detection are very useful for gene cloning, MAS breeding, and trait improvement; however, only a few studies have been performed using SSR markers in adzuki bean (Isemura et al., 2007; Kaga et al., 2008). We identified 26 genomic QTLs associated with agronomic traits and gene loci of four qualitative traits using a high-density SNP genetic map created via RAD sequencing. In this study, F_2 segregating populations were derived using a wide cross between a cultivated adzuki bean Ass001 (*V. angularis* var. *angularis*) and a wild adzuki bean CWA108 (*V. angularis* var. *nipponensis*). The seeds of the wild adzuki bean were in dormancy. The seed number of F_2 individuals was limited. Each of the $F_{2:3}$ lines were planted in

two rows in 2014. Each of the $F_{3:4}$ lines were planted in the same manner as the $F_{2:3}$ lines in 2015. From the center of the rows, 10 representational plants per line and 10 plants of each parent were selected to evaluate each trait based on the mean value. Isemura et al. (2007) planted $BC_1F_{1:2}$ lines and $F_{2:3}$ lines, each consisting of 10 individuals per line in the field, in June 2003 and 2004 at NIAS, and the lines were evaluated based on the mean values for each trait per line and used for QTL mapping. Kaga et al. (2008) used the same planting model and method for QTL identification in adzuki bean. Wu et al. (2013) mapped the QTL of grain shape and size in wheat RILs using 10 representational plants that were selected from the center of two rows to gain phenotypic data. The phenotypic trait data were reliable in this study.



Control of flowering time is critical when pulse crops have to adapt to different ecogeographic environments and photoperiods. In the soybean, a representative warm-season short-day legume, at least 10 *E* loci (*E1–E9*) and *J* genes controlling flowering time and maturity have been described (Weller and Ortega, 2015; Cao et al., 2017). In the adzuki bean, a first-flowering-day (FLD) QTL has been described in linkage group 4 (Isemura et al., 2007), and a major relevant QTL (*Fld3.4a.1*, PEV 43.7%) in LG4a; nine FLD and pod maturity QTLs were found to be present on LG2, LG3, LG4a, LG5, LG6a, and LG11 (Kaga et al., 2008). In the present study, we detected 10 QTLs controlling flowering time and pod maturity; the QTLs were on chromosomes 4, 7, and 10. The very relevant *VaFld4.1* and *VaFld4.2* loci were on chromosome 4, and the candidate phytochrome E gene *VaPhyE* co-segregated with SNP markers. Phytochrome-A-encoding genes *E3* and *E4* have been identified in the soybean (Liu et al., 2008; Watanabe et al., 2009). *GmE3* and *GmE4* are *phyA*-type (phytochrome A) photoreceptor that plays an important role in the detection of far-red and red light in the soybean, and is associated with early flowering and maturity. *E4*, controls flowering when the LDs are in the low- and far-red regions (Cober and Voldeng, 2001). *E3* functions across a wide range of latitudes, whereas allelic differences in *E4* are detected only at high latitudes (Lu et al., 2015). *GmPHYA1* and *GmPHYA2* (*E4*) may play redundant roles in photomorphogenesis (de-etiolation response and flowering under low R:FR conditions; Liu et al., 2008). Soybean has eight PHY loci including four PHYA, two PHYB, and two PHYE loci. Phylogenetic analysis of soybean suggests that the four paralogous PHY pairs separated (by genomic duplication) at least 13 million years ago, and that the four PHYA copies are remnants of two rounds of such duplication (ca. 58 and 13 million years ago) (Wu et al., 2013).

Adzuki bean, like most sequenced legumes, did not experience this *Glycine max*-specific whole genome duplication event (13 million years ago). In this study, we found that there are four PHY loci including two PHYA, a PHYB, and a PHYE loci in adzuki bean, common bean, mung bean, chickpea, pigeonpea, and *Arachis duranensis*, but a PHYA in *Medicago truncatula*. This result confirms that two rounds of whole genome duplication events led to four PHYA and two PHYE copies in soybean. The PHY family includes the A, B, C, D, and E subfamilies; PHYC and PHYD were not found in the adzuki bean, soybean, common bean, chickpea, pigeonpea, mung bean, *Arachis duranensis*, or *M. truncatula* genomes. PHYC and PHYD may have been lost during the evolution of legumes. PHYE was absent in monocotyledon *Oryza sativa* and *Zea mays* (Table S4). *Dt2*, a dominant MADS-domain gene of the *AP1/SQUA* subfamily, has been cloned from the soybean. *Dt2* represses *Dt1* expression in the shoot apical meristem, promoting the early conversion of meristem into reproductive inflorescences; the gene thus promotes semi-determinate stem growth and early flowering (Ping et al., 2014). In this study, 28 AGAMOUS-like MADS-box genes were found in the adzuki bean genome for photoperiodic flowering (Table S6). Kim et al. (2014) sequenced a *de novo* assembly of adzuki bean transcript sequences and subjected them to a BLASTP search to identify putative homologs of the 84 *Arabidopsis* genes involved in the circadian clock and photoperiodic flowering pathway. Eleven homologs of the AGAMOUS-like MADS-box transcription factor were detected in the adzuki bean. Follow-up studies of candidate gene validation are being planned.

AP2 genes are representative of the large AP2/EREBP gene family, and are both necessary and sufficient for flower development (Jofuku et al., 1994), stem cell maintenance (Wurschum et al., 2005), seed development, determination

TABLE 4 | QTLs mapped of agronomic traits in F₂, F₃, and F₄ populations.

Trait ^a	QTL	Chr	QTL peak (cM)	Flanking marker ^b	SNP (P1/P2) ^c	Genetic distance between loci and trait cM	Left flanking marker (cM)	Right flanking marker (cM)	Genomic position ^d	LOD	Expl%
BRN	VaBrn4.1 (F ₂ ,F ₃)	4	18.455	s690-144110,s856-165880	(T/G)	0.0593	17.862	18.455	2728770-2973972	5.91	20.4
	VaBrn4.2 (F ₃)	4	17.862	s165-116310	(G/A)	0.974	16.888	16.888	2466844	4.55	20.8
ST1L_ST10l	VaST1-10l9.1 (F ₂)	9	57.735	s249-3204596	(C/A)	0	57.735	57.735	28712717	4.92	17.3
SDNPPD	VaSnpp3.1 (F ₂)	3	16.847	s624-5207,s624-5245	(A/C)	0.0120	16.835	16.847	5754076-5754114	4.89	17.3
	VaSnpp4.1 (F ₂)	4	19.37	s80-1776704	(G/A)	0	19.37	19.37	5393066	3.98	11.8
SD100WT	VaSd100wt11.1 (F ₂)	11	57.237	s268-1701932,s268-1618129	(C/T)	0.0420	57.195	57.237	31820598-31904338	6.00	20.7
	VaSd100wt1.1 (F ₂ ,F ₃)	1	92.683	s244-1339845,s244-641945	(A/G)	2.143,13.457	90.54	106.14	32224913-32922813	4.34	12.2
	VaSd100wt1.2 (F ₃)	1	88.54	s168-1117751	(G/A)	19.429	88.54	88.54	30134755	4.16	19.2
PDDM25	VaSd100wt1.3(F ₂)	1	90.683	s244-641945,s384-495870	(G/A)	0.5457	90.683	96.14	32922813-36773639	4.06	11.5
	VaFId4.1 (F ₂)	4	17.862	s690-144110	(A/C)	0	17.862	17.862	2728770	20.00	55.4
PDDM50	VaFId4.1 (F ₂)	4	17.862	s690-144110	(A/C)	0	17.862	17.862	2728770	25.5	64.9
PDDM75	VaGp4.1 (F ₂ ,F ₃)	4	20.113	s80-1776704,s80-1939896	(C/T)	0.7431	19.37	21.113	5556258	14.46	47
	VaFId4.1 (F ₂ ,F ₄ ,F ₂)	4	17.862	s165-116310,s690-144110	(A/C)	0.9740	16.888	17.862	2466844-2728770	32.29	71.3
FLD	VaFId4.2 (F ₂ ,F ₂)	4	17.862	s690-144110,s856-165880	(A/C)	0.0593	17.862	18.455	2728770-2973972	32.29	25.9
	VaFId4.3(F ₄ ,F ₄)	4	16.888	s58-238422,s165-116310	(G/A)	5.4490	12.407	16.888	238422-2466844	6.19	30.3
	VaFId7.1 (F ₂)	7	45.625	s57-1742194	(G/T)	0	45.625	45.462	8738088	4.00	4.1
	VaFId4.1 (F ₂ ,F ₂)	4	17.862	s165-116310, s690-144110	(A/C)	0	16.888	17.862	2466844-2728770	33.96	73.1
FLD50	VbFId4.1 (F ₃ ,F ₄)	4	22.272	s80-940518,s80-1199058	(C/A)	0.6330	21.639	22.272	4556880-4815420	4.69	24.2
	VaFId10.1 (F ₂)	10	62.007	s20-880363, s356-1554658	(A/C)	0	62.007	62.007	35805580-37408870	4.05	3.9
	VaFId10.2(F ₂)	10	62.738	s356-1554658	(T/G)	0	62.735	62.735	37408870	4.00	3.9
LFMW	VaLfmw2.1 (F ₄)	2	50.877	s536-414880	(C/G)	0	50.877	50.877	22751890	4.1	21.2
	VaSdt4.1 (F ₂)	4	9.641	s58-1489334,s58-345785	(G/T)	9.6410	0	9.641	345785,1489334	2.84	9.11

The Permutation Test is used to calculate the significant threshold for subsequent Interval mapping and MQM mapping; "Expl(%)" denotes the percentage of the variance (Van Ooijen, 2009); "QTL Peak" presents a loci with the largest LOD value. This loci identifies the most likely location that is located within a gene or near a gene affected by a relative trait.

^aQTL names: BRN, number of branches on the main stem; ST1L_ST10l, stem internode length (first to tenth); SDNPPD, seed number per Pod; SD100WT, 100-seed weight of pod; PDDM25, days to maturity of 25% pod; PDDM50, days to maturity of 50% pod; PDDM75, days to maturity of 75% pod; FLD, days to first flowering; FLD50, days of 50% flowering.

^bThe left flanking marker, the right flanking marker: if there is only one marker at a site, only one marker met our criterion (LOD > threshold value of the Permutation Test).

^cP1 is the abbreviation of parent 1 which is one of the parents of the QTL group, P2 represents another parent.

^dGenomic position: genomic position of the left flanking marker, genomic position of the right flanking marker.

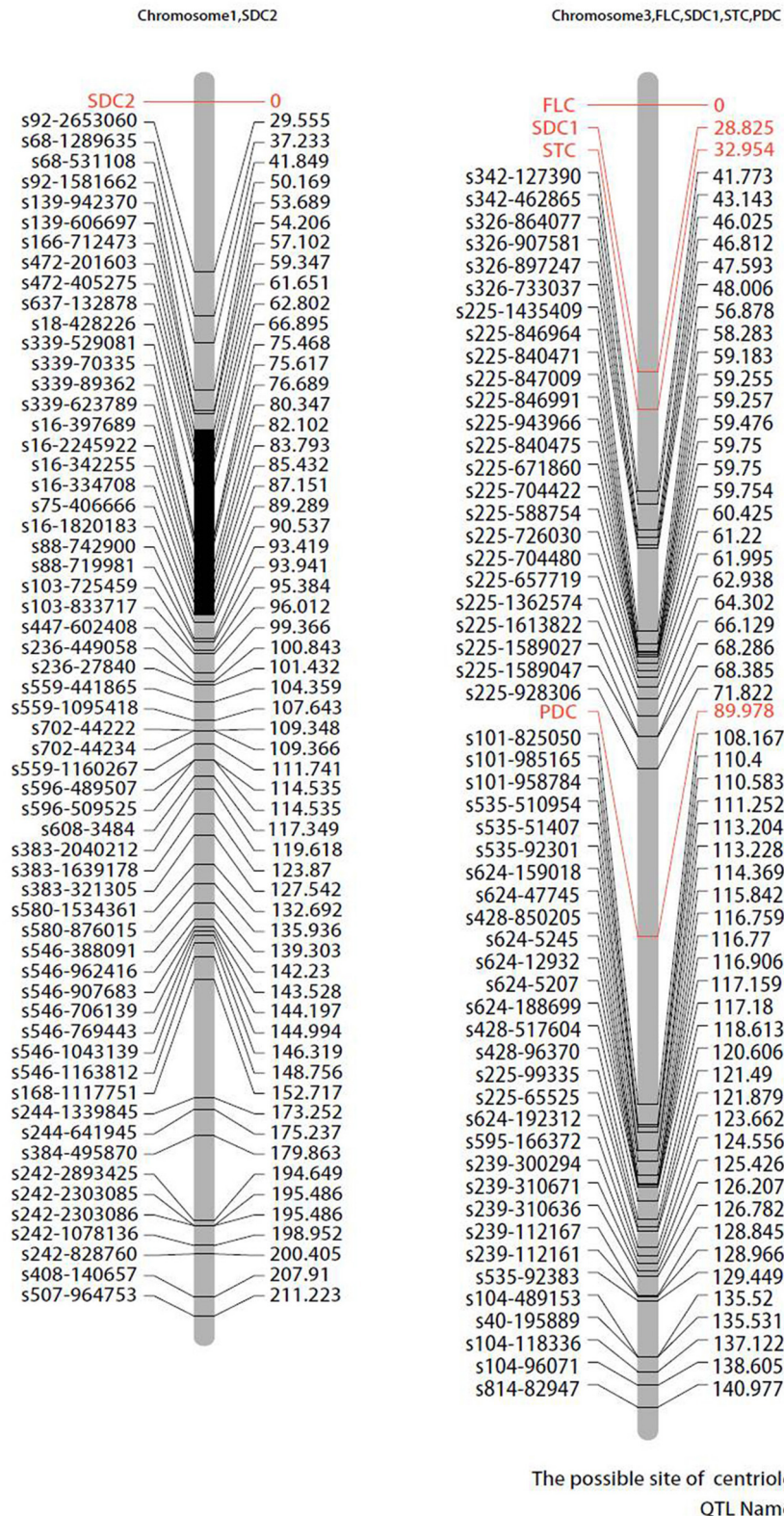


FIGURE 5 | Mapping of qualitative traits on the map of the population derived from the cross between cultivated and wild adzuki bean. SDC, seed coat color; FLC, flower color; STC, stem color; PDC, pod color.

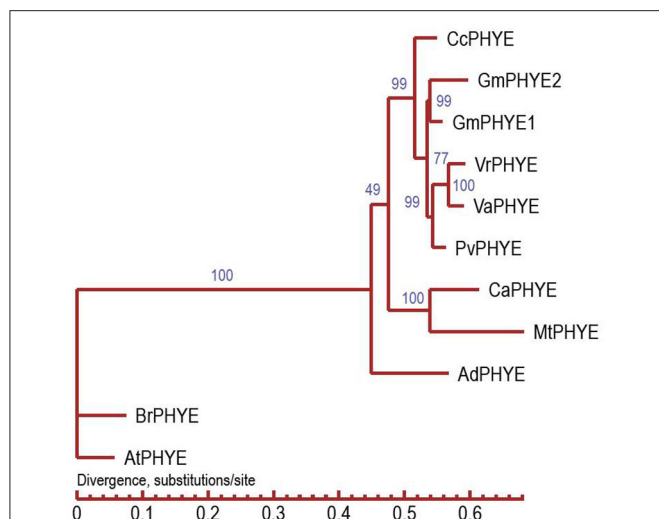


FIGURE 6 | Phylogenetic tree analysis of adzuki bean PHYE. CcPHYE (*Cajanus cajan*), GmPHYE1 and GmPHYE2 (*Glycine max*), VrPHYE (*Vigna radiata*), VaPHYE (*Vigna angularis*), PvPHYE (*Phaseolus vulgaris*), CaPHYE (*Cicer arietinum*), MtPHYE (*Medicago truncatula*), AdPHYE (*Arachis duranensis*), BrPHYE (*Brassica rapa*), and AtPHYE (*Arabidopsis thaliana*).

of seed size, and seed weight (Jofuku et al., 2005); several AP2-encoding proteins repress flowering, redundantly affecting flowering time (Zhu and Helliwell, 2011). Four *VaAP2* were identified in QTL mapping intervals of seed weight, leaf size, and seed thickness in adzuki bean. We concluded that these *VaAP2* might function in regulating organ size, but their function still requires validation. If leaf width candidate gene *VaAP2-l2* is involved in or regulates seed size, this should be analyzed in future studies.

The black seed coat color *VaScB* and red seed coat color *VaScR* genes mapped to chromosomes 3 and 1, respectively. Kaga et al. (2008) mapped the red seed coat locus near the SSR marker CEDG053 on linkage group 1 (LG1). Genes that control red or tan seed coat colors, and black seed mottling have been mapped to linkage groups 1 and 4, respectively (Isemura et al., 2007). A minor QTL *OLB2* explaining 6% of the total variance in redness is located on LG1 (Horiuchi et al., 2015). *VaScR* will be finely mapped by reference to the closest SNP and SSR markers in follow-up studies. *UGT* is a member of the flavonoid metabolic pathway, and is involved in anthocyanin biosynthesis. The enzymes O-methyltransferase, rhamnosyl transferase, and UDP flavonoid glucosyl transferase together synthesize anthocyanins from 3-OH-anthocyanidins. UDP glucosyl transferase is essential for anthocyanin synthesis.

REFERENCES

- Amarowicz, R., Estrella, I., Hernández, T., and Troszyńska, A. (2008). Antioxidant activity of extract of adzuki bean and its fractions. *J. Food Lipids* 15, 119–136. doi: 10.1111/j.1745-4522.2007.00106.x

The enzyme may be involved in delphinidin synthesis; this pigment is blue–black in color, and may be responsible for the difference between black (pigment present) and brown (pigment absent) seed coat colors (Lepiniec et al., 2006). We designed primers and found a new SNP cosegregative marker with *VaScB*. The proposed function of *VaUGT* for synthesis of a black seed coat needs to be verified experimentally. A *UGT78K1* (UDP-glucose: flavonoid-3-O-glycosyltransferase) gene, imparting a black seed coat, has been isolated from the black (iRT) soybean (Kovinich et al., 2010). *VaUGT* belongs to 73C3-like UDP-glycosyl transferase, and is different from *GmUGT78K1*; its function will need to be further verified.

AUTHOR CONTRIBUTIONS

PW designed and managed the project and wrote the manuscript. YL prepared DNA samples for RAD sequencing, and engaged in data investigation and analysis. YL, KY, WY, and JJ performed QTL analyses. YL and KY identified candidate flowering time genes. BZ, YSL, and TW participated in trait investigations, data analysis, and field experiments. LC and ZY mapped qualitative traits and identified relevant candidate genes. CC analyzed the phylogenetic tree. KY contributed helpful suggestions.

ACKNOWLEDGMENTS

The National Natural Science Foundation of China provided financial support (grant nos. 31371694, 31272238, and 31571734).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00840/full#supplementary-material>

Table S1 | Phenotype Data of Parents and F₂, F₃ and F₄ Populations.

Table S2 | Comparison of sequences assembly between parents and 150 individuals of F₂ by RAD-sequencing.

Table S3 | Phytochrome genes in the adzuki bean and comparison to sequenced legume and other plant genomes.

Table S4 | *Agamous-like MADS-box* genes in the adzuki bean and comparison to sequenced legume and other plant genomes.

Table S5 | *AP2/EREBF* genes in the adzuki bean and comparison to sequenced legume and other plant genomes.

Table S6 | UDP flavonoid glycosyl transferase (*UGT*) genes in the adzuki bean and comparison to sequenced legume and other plant genomes.

Figure S1 | Phylogenetic tree analysis of adzuki bean *UGT*-like genes.

- Aoyama, S., Onishi, K., and Kato, K. (2011). The genetically unstable dwarf locus in azuki bean (*Vigna angularis* Willd. Ohwi & Ohashi). *J. Hered.* 102, 604–609. doi: 10.1093/jhered/esr068
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376. doi: 10.1371/journal.pone.0003376

- Cao, D., Takeshima, R., Zhao, C., Liu, B., Jun, A., and Kong, F. (2017). Molecular mechanisms of flowering under long days and stem growth habit in soybean. *J. Exp. Bot.* 68, 1873–1884. doi: 10.1093/jxb/erw394
- Cheng, X. Z., and Tian, J. (2009). "Status and future perspectives of Vigna (mungbean and azuki bean) production and research in China," in *The 14th National Institute of Agrobiological Sciences, Japan, International Workshop on Genetic Resources, Genetics Resources and Comparative Genomics of Legumes* (Tsukuba: Glycine and Vigna), 83–86.
- Cheng, X. Z., Wang, S. H., Wang, L. X., Wang, S. M., Zong, X. X., Wang, P. Z., et al. (2012). *Descriptors and Data Standard for Adzuki Bean [Vigna angularis (Willd.) Ohwi & Ohashi]*, eds X. Z. Cheng, S. H. Wang, and L. X. Wang. Beijing: China Agriculture Press.
- Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Cober, E. R., and Voldeng, H. D. (2001). Low R: FR light quality delays flowering of E7E7 soybean lines. *Crop Sci.* 41, 1823–1826. doi: 10.2135/cropsci2001.1823
- de Folter, S., Immink, R. G. H., Kieffer, M., Pařenicová, L., Henz, S. R., Weigel, D., et al. (2005). Comprehensive interaction map of the Arabidopsis MADS-box transcription factors. *Plant Cell* 17, 1424–1433. doi: 10.1105/tpc.105.031831
- Grattapaglia, D., and Sederoff, R. (1994). Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. *Genetics* 137, 1121–1137.
- Han, A., Kaga, A., Isemura, T., Wang, X., Tomooka, N., and Vaughan, D. A. (2005). A genetic linkage map for azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. *Theor. Appl. Genet.* 111, 1278–1287. doi: 10.1007/s00122-005-0046-8
- Horiuchi, Y., Yamamoto, H., Ogura, R., Shimoda, N., Sato, H., and Kato, K. (2015). Genetic analysis and molecular mapping of genes controlling seed coat colour in adzuki bean (*Vigna angularis*). *Euphytica* 206, 609–617. doi: 10.1007/s10681-015-1461-9
- Isemura, T., Kaga, A., Konishi, S., Ando, T., Tomooka, N., Han, O. K., et al. (2007). Genome dissection of traits related to domestication in azuki bean (*Vigna angularis*) and comparison with other warm-season legumes. *Ann. Bot.* 100, 1053–1071. doi: 10.1093/aob/mcm155
- Jansen, R. C. (1993). Interval mapping of multiple quantitative trait loci. *Genetics* 135, 205–211.
- Jofuku, K. D., Den Boer, B., Van Montagu, M., and Okamuro, J. K. (1994). Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6, 1211–1225. doi: 10.1105/tpc.6.9.1211
- Jofuku, K. D., Omidyar, P. K., Gee, Z., and Okamuro, J. K. (2005). Control of seed mass and seed yield by the floral homeotic gene APETALA2. *Proc. Natl. Acad. Sci. U.S.A.* 102, 3117–3122. doi: 10.1073/pnas.0409893102
- Kaga, A., Isemura, T., Tomooka, N., and Vaughan, D. A. (2008). The genetics of domestication of the azuki bean (*Vigna angularis*). *Genetics* 178, 1013–1036. doi: 10.1534/genetics.107.078451
- Kim, M. Y., Yang, J. K., and Lee, S. H. (2014). Circadian clock and photoperiodic flowering genes in adzuki bean (*Vigna angularis* [Willd.] Ohwi & H. Ohashi). *Plant Genet. Res.* 12, S49–S53. doi: 10.1017/s1479262114000252
- Kosambi, D. D. (1994). The estimation of map distance from recombination values. *Ann. Eugen.* 12, 12172–12175.
- Kovinich, N., Saleem, A., Arnason, J. T., and Miki, B. (2010). Functional characterization of a UDP-glucose:flavonoid 3-O-glucosyltransferase from the seed coat of black soybean (*Glycine max* (L.) Merr.). *Phytochem.* 71, 1253–1263. doi: 10.1016/j.phytochem.2010.05.009
- Kramer, C., Soltani, N., Robinson, D. E., Swanton, C. J., and Sikkema, P. H. (2012). Control of volunteer adzuki bean in soybean. *Agri. Sci.* 3, 501–509. doi: 10.4236/as.2012.34059
- Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., and Nesi, N. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252
- Lin, R. F. (2002). *Minor Grain Crops in China*, eds Y. Chai, Q. Liao, and S. W. Sun. Beijing: China Agricultural Science and Technology Press.
- Liu, B., Kanazawa, A., Matsumura, H., Takahashi, R., Harada, K., and Abe, J. (2008). Genetic redundancy in soybean photoresponses associated with duplication of the phytochrome A gene. *Genetics* 180, 995–1007. doi: 10.1534/genetics.108.092742
- Liu, L., Bestel, S., Shi, J., Song, Y., and Chen, X. (2013). Paleolithic human exploitation of plant foods during the last glacial maximum in North China. *Proc. Natl. Acad. Sci. U.S.A.* 110, 5380–5385. doi: 10.1073/pnas.1217864110
- Lu, S., Li, Y., and Wang, J. (2015). QTL mapping for flowering time in different latitude in soybean. *Euphytica* 206, 725–736. doi: 10.1007/s10681-015-1501-5
- Ping, J., Liu, Y., Sun, L., Zhao, M., Li, Y., She, M., et al. (2014). Dt2 is a gain-of-function MADS-domain factor gene that specifies semi-determinacy in soybean. *Plant Cell* 26, 2831–2842. doi: 10.1105/tpc.114.126938
- Tomooka, N., Vaughan, D., Moss, H., and Maxted, N. (2002). *The Asian Vigna: Genus Vigna Subgenus Ceratotropis Genetic Resources*, ed N. Tomooka. Dordrecht: Kluwer Academic Publishers.
- Van Ooijen, J. W. (2006). *JoinMap 4: Software for the Calculation of Genetic Linkage Maps in Experimental populations*, ed B. V. Kyazma. Wageningen.
- Van Ooijen, J. W. (2009). "MapQTL6: Software for the mapping of quantitative trait loci in experimental populations," in *Plant & Animal Genome XVII Conference*, ed B. V. Kyazma (Wageningen).
- Wang, X. W., Kaga, A., Tomooka, N., and Vaughan, D. A. (2004). The development of SSR markers by a new method in plants and their application to gene flow studies in azuki bean [*Vigna angularis* (Willd.) Ohwi & Ohashi]. *Theor. Appl. Genet.* 109, 352–360. doi: 10.1007/s00122-004-1634-8
- Watanabe, S., Hideshima, R., Xia, Z., Tsubokura, Y., Sato, S., Nakamoto, Y., et al. (2009). Map-based cloning of the gene associated with the soybean maturity locus E3. *Genetics* 182, 1251–1262. doi: 10.1534/genetics.108.098772
- Weller, J. L., and Ortega, R. (2015). Genetic control of flowering time in legumes. *Front. Plant Sci.* 6:207. doi: 10.3389/fpls.2015.00207
- Wu, F. Q., Fan, C. M., Zhang, X. M., and Fu, Y. F. (2013). The phytochrome gene family in soybean and a dominant negative effect of a soybean PHYA transgene on endogenous Arabidopsis PHYA. *Plant Cell Rep.* 32, 1879–1890. doi: 10.1007/s00299-013-1500-8
- Wurschum, T., Gross-Hardt, R., and Laux, T. (2005). APETALA2 regulates the stem cell niche in the Arabidopsis shoot meristem. *Plant Cell* 18, 295–307. doi: 10.1105/tpc.105.038398
- Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., et al. (2015). Genome sequencing of adzuki bean (*Vigna angularis*) provides insight into high starch and low fat accumulation and domestication. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13213–13218. doi: 10.1073/pnas.1420949112
- Yao, Y., Cheng, X., Wang, S., Wang, L., and Ren, G. (2012). Influence of altitudinal variation on the antioxidant and antidiabetic potential of azuki bean (*Vigna angularis*). *Int. J. Food Sci. Nutr.* 63, 117–124. doi: 10.3109/09637486.2011.604629
- Zhu, Q.-H., and Helliwell, C. A. (2011). Regulation of flowering time and floral patterning by miR172. *J. Exp. Bot.* 62, 487–495. doi: 10.1093/jxb/erq295

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Li, Yang, Yang, Chu, Chen, Zhao, Li, Jian, Yin, Wang and Wan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A High-Resolution InDel (Insertion–Deletion) Markers-Anchored Consensus Genetic Map Identifies Major QTLs Governing Pod Number and Seed Yield in Chickpea

Rishi Srivastava^{1†}, Mohar Singh^{2†}, Deepak Bajaj¹ and Swarup K. Parida^{1*}

¹ National Institute of Plant Genome Research, New Delhi, India, ² National Bureau of Plant Genetic Resources Regional Station, Shimla, India

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Fangsen Xu,
Huazhong Agricultural University,
China
Milind Ratnaparkhe,
ICAR-Indian Institute of Soybean
Research, India

*Correspondence:

Swarup K. Parida
swarup@nipgr.ac.in;
swarupdbt@gmail.com

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 14 June 2016

Accepted: 29 August 2016

Published: 16 September 2016

Citation:

Srivastava R, Singh M, Bajaj D and
Parida SK (2016) A High-Resolution
InDel (Insertion–Deletion)
Markers-Anchored Consensus
Genetic Map Identifies Major QTLs
Governing Pod Number and Seed
Yield in Chickpea.
Front. Plant Sci. 7:1362.
doi: 10.3389/fpls.2016.01362

Development and large-scale genotyping of user-friendly informative genome/gene-derived InDel markers in natural and mapping populations is vital for accelerating genomics-assisted breeding applications of chickpea with minimal resource expenses. The present investigation employed a high-throughput whole genome next-generation resequencing strategy in low and high pod number parental accessions and homozygous individuals constituting the bulks from each of two inter-specific mapping populations [(Pusa 1103 × ILWC 46) and (Pusa 256 × ILWC 46)] to develop non-erroneous InDel markers at a genome-wide scale. Comparing these high-quality genomic sequences, 82,360 InDel markers with reference to *kabuli* genome and 13,891 InDel markers exhibiting differentiation between low and high pod number parental accessions and bulks of aforementioned mapping populations were developed. These informative markers were structurally and functionally annotated in diverse coding and non-coding sequence components of genome/genes of *kabuli* chickpea. The functional significance of regulatory and coding (frameshift and large-effect mutations) InDel markers for establishing marker-trait linkages through association/genetic mapping was apparent. The markers detected a greater amplification (97%) and intra-specific polymorphic potential (58–87%) among a diverse panel of cultivated *desi*, *kabuli*, and wild accessions even by using a simpler cost-efficient agarose gel-based assay implicating their utility in large-scale genetic analysis especially in domesticated chickpea with narrow genetic base. Two high-density inter-specific genetic linkage maps generated using aforesaid mapping populations were integrated to construct a consensus 1479 InDel markers-anchored high-resolution (inter-marker distance: 0.66 cM) genetic map for efficient molecular mapping of major QTLs governing pod number and seed yield per plant in chickpea. Utilizing these high-density genetic maps as anchors, three major genomic regions harboring each of pod number and seed yield robust QTLs (15–28% phenotypic variation explained) were identified on chromosomes 2, 4, and 6. The integration of genetic and physical maps

at these QTLs mapped on chromosomes scaled-down the long major QTL intervals into high-resolution short pod number and seed yield robust QTL physical intervals (0.89–2.94 Mb) which were essentially got validated in multiple genetic backgrounds of two chickpea mapping populations. The genome-wide InDel markers including natural allelic variants and genomic loci/genes delineated at major six especially in one colocalized novel congruent robust pod number and seed yield robust QTLs mapped on a high-density consensus genetic map were found most promising in chickpea. These functionally relevant molecular tags can drive marker-assisted genetic enhancement to develop high-yielding cultivars with increased seed/pod number and yield in chickpea.

Keywords: chickpea, genetic map, InDel, pod number, QTL, seed yield

INTRODUCTION

Insertion–deletions (InDels) are the preferred ideal sequence-based marker for driving genomics-assisted breeding applications in multiple crop plants. This is due to its myriad of desirable inherent genetic attributes in conjunction with other genetic markers like simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs; Li et al., 2014; Moghaddam et al., 2014; Wang et al., 2014; Das et al., 2015). The available draft whole genome sequences and next-generation sequencing (NGS) genome/transcriptome resequences of diverse *desi*, *kabuli*, and wild accessions are found expedient to develop genome-wide including gene-derived InDel markers *in-silico* with minimal resource expenses in chickpea (Agarwal et al., 2012; Hiremath et al., 2012; Jain et al., 2013; Varshney et al., 2013; Deokar et al., 2014; Kudapa et al., 2014). Recently, the advantages of InDel markers structurally/functionally annotated at a whole genome and gene level are well-demonstrated in various large-scale genotyping applications of chickpea (Das et al., 2015). Essentially, this involves understanding the genetic diversity and phylogeny among cultivated *desi*, *kabuli*, and wild accessions, constructing high-density genetic linkage maps and molecular mapping of major QTLs governing various important agronomic traits like flowering and maturity time in chickpea (Das et al., 2015). Despite these efforts, hitherto none of the informative InDel markers tightly linked to the major genes/QTLs regulating a/biotic stress tolerance and yield component traits has been validated in multiple genetic backgrounds and delineated by genetic/association mapping to be exploited for marker-assisted genetic improvement of chickpea. The narrow genetic base, including low marker genetic polymorphism especially among mapping and natural populations coupled with inadequate accessibility of high-density genetic linkage maps are the major bottlenecks in identification and fine mapping/map-based cloning of trait-associated QTLs in chickpea. In these perspectives, development and high-throughput genotyping of numerous genome-wide informative InDel markers in mapping populations and natural germplasm lines (association panel) for generating high-density genetic linkage maps, high-resolution QTL mapping (fine mapping/positional cloning), and genetic association analysis are essential in chickpea. This will essentially assist us to delineate functionally relevant genes/QTLs and natural allelic variants governing vital agronomic traits for

genomics-assisted crop improvement of chickpea with narrow genetic base.

In light of the above, the present study has made efforts to develop large-scale high-quality InDel markers at a genome-wide scale by employing a high-throughput NGS resequencing strategy in low and high pod number parental accessions and bulks (homozygous mapping individuals representing extreme pod number phenotypic trait values) constituted from two F₅ mapping populations of *Cicer arietinum desi* and *Cicer reticulatum* wild inter-specific crosses. These genome-wide InDel markers were further utilized to detect potential of intra-/inter-specific polymorphism among cultivated (*desi* and *kabuli*) and wild chickpea accessions. The significance of InDel markers was further assessed to construct a high-density consensus inter-specific genetic linkage map and for efficient high-resolution molecular mapping of major genes/QTLs regulating vital agronomic traits including pod number and seed yield per plant with a prime objective of accelerating marker-assisted genetic enhancement in chickpea.

MATERIALS AND METHODS

Development of Whole Genome Resequencing-Based InDel Markers

Two inter-specific F₅ mapping populations [(*desi* Pusa 1103 × wild ILWC 46, 102 individuals) and (*desi* Pusa 256 × ILWC 46, 98 individuals)] derived from inter-crosses between *C. arietinum desi* and *C. reticulatum* wild accessions were developed. To identify more robust InDels, the high-quality mappable pair-end (100-bp read length), and normalized NGS genome resequencing data of parental accessions were acquired from the afore-mentioned two mapping populations individually as per our previous study (Das et al., 2016). Like-wise, genome resequences from 10 of each low and high pod number homozygous mapping individuals (representing two utmost ends of pod number normal frequency distribution curve) constituting the low pod number bulk (LPNB) and high pod number bulk (HPNB), respectively were obtained. The sequence reads of parental accessions as well as bulks (HPNB and LPNB) obtained from two inter-specific mapping populations were mapped onto the chromosome pseudomolecules and unanchored scaffolds of reference *kabuli* (CDC Frontier)

genome (Varshney et al., 2013). Subsequently, high-quality (minimal false-positive) InDels among mapping parents and bulks/individuals were detected following Das et al. (2015). To develop genome-wide InDel markers, forward and reverse primers from the CDC Frontier *kabuli* genomic sequences flanking the InDels were designed using Primer3 interface module of MISA (<http://pgrc.ipk-gatersleben.de/misa/primer3.html>). The developed InDel markers were structurally and functionally annotated with respect to *kabuli* genome as per Das et al. (2015) and Kujur et al. (2015a). The KOG (eukaryotic orthologous groups of proteins, <ftp://ftp.ncbi.nih.gov/pub/COG/KOG>) and transcription factor (TF) gene-based functional annotation of InDel markers were performed in accordance with Das et al. (2015) and Kujur et al. (2015a).

Experimental Validation and Polymorphic Potential of InDel Markers

To evaluate the amplification and polymorphic potential of InDel markers developed from two mapping populations, the InDel markers exhibiting ≥ 4 bp *in-silico* fragment length polymorphism between parental accessions and bulks (LPNB/HPNB) of two inter-specific mapping populations were selected. These markers were PCR amplified and genotyped by agarose gel- and PCR amplicon resequencing-based assays using the genomic DNA of 24 cultivated and wild chickpea accessions. This included three mapping parental accessions (Pusa 256, Pusa 1103, and ILWC 46) from which the InDel markers were originally identified, and 21 additional *desi* (4) and *kabuli* (3) and wild (14) chickpea accessions. The genotyping data of experimentally validated InDel markers were utilized to measure the average polymorphic alleles per marker, percent polymorphism and polymorphism information content (PIC) among chickpea accessions employing PowerMarker v3.51 (<http://statgen.ncsu.edu/powermarker>).

Genetic Linkage Map Construction

The InDel markers exhibiting polymorphism between parental accessions (Pusa 1103 vs. ILWC 46 and Pusa 256 vs. ILWC 46) were PCR amplified and genotyped using 102 and 98 individuals derived from two F₅ inter-specific mapping populations of Pusa 1103 \times ILWC 46 and Pusa 256 \times ILWC 46, respectively using agarose gel- and PCR amplicon resequencing-based assays. The marker genotyping data were analyzed with the χ^2 -test ($P < 0.05$) to evaluate their goodness-of-fit to the expected Mendelian 1:1 segregation ratio. The JoinMap 4.1 (<http://www.kyazma.nl/index.php/mc.JoinMap>) at a higher logarithm of odds (LOD) threshold (4.0–8.0) with Kosambi mapping function was used to measure the linkage analysis among InDel markers. The InDel markers were assigned into defined LGs (linkage groups; designated as LG1 to LG8)/chromosomes of two inter-specific genetic maps according to their centiMorgan (cM) genetic distances and corresponding marker physical positions (bp) on the chromosomes. A consensus high-density genetic linkage map derived from two inter-specific genetic maps was constructed using JoinMap 4.1 (following Bohra et al., 2012; Varshney et al., 2014) and visualized using Circos as per Kujur et al. (2015a).

QTL Mapping

The mapping individuals and parental accessions of two F₅ inter-specific mapping populations [(Pusa 1103 \times ILWC 46) and (Pusa 256 \times ILWC 46)] were grown in the field as per random complete block design (RCBD) with at least two replications and phenotyped at two diverse geographical locations of India (CSKHPKV, Palampur: latitude 32.1°N and longitude 76.5°E and NBPGR, New Delhi: 28.6°N and 77.2°E) for two consecutive years (2013 and 2014). For precise phenotyping, 10–15 representative plants from each mapping individual and parental accession of both mapping populations were selected to estimate the pod number and seed yield (g) per plant. The pod number (PN) was measured by counting the average number of fully developed pods per plant at maturity stage whereas seed yield per plant (SYP) was estimated by taking average weight (g) of fully matured dried seeds (at 10% moisture content) harvested from all representative plants belonging to each mapping individual and parental accession of aforesaid both populations. The inheritance pattern of PN and SYP based on diverse statistical parameters including mean, standard deviation, coefficient of variation (CV), broad-sense heritability (H^2), Pearson's correlation coefficient and frequency distribution was measured in both mapping populations individually following Bajaj et al. (2015a) and Das et al. (2016).

For molecular mapping of major PN and SYP QTLs, the genotyping data of InDel markers genetically mapped on two individual and/or a consensus high-density chickpea genetic linkage map (comprising of eight LGs/chromosomes) was integrated with PN and SYP field phenotypic data of mapping individuals and parental accessions using a composite interval mapping (CIM) function of MapQTL 6 (Van Ooijen, 2009) as per Varshney et al. (2014) and Das et al. (2015). The LOD threshold score > 4.0 with 1000 permutations at a $p < 0.05$ significance was used as major criteria in CIM for QTL mapping. The phenotypic variation explained (PVE) and additive effect specified by each major PN and SYP QTL at a significant LOD were measured in accordance with Bajaj et al. (2015a).

RESULTS

Whole-Genome Resequencing of Low and High Pod Number Parental Accessions and Homozygous Bulks from Mapping Populations

We generated on an average 81.5 million high-quality sequence reads (with a ~ 11.6 -fold sequencing depth coverage) by high-throughput whole-genome NGS resequencing of two low and high pod number parental accessions as well as bulks (LPNB and HPNB) from each of two inter-specific mapping populations [(Pusa 1103 \times ILWC 46) and (Pusa 256 \times ILWC 46)]. To reduce the potential biasness of read depth in the examined samples, the high-quality uniquely mapped non-redundant sequence reads (69% mean coverage on *kabuli* reference genome) generated from parental accessions and bulks (LPNB and HPNB) of two mapping populations were normalized based on depth of read coverage. This analysis

revealed ~11.6-fold average sequencing depth coverage and 64.1% (474.2 Mb) mean genome coverage (%) of *kabuli* chickpea (with an estimated genome size of ~740 Mb) in mapping parents and bulks. All these sequencing data were submitted to NCBI-sequence read archive (SRA) database (<http://www.ncbi.nlm.nih.gov/sra>) with accession number SRR2228974 for unrestricted public access. The normalized sequence reads of parents and bulks from each of two inter-specific mapping populations were compared individually with reference *kabuli* genomic sequences (pseudomolecules and unanchored scaffolds) to discover the high-quality non-erroneous InDels.

Development of Genome-Wide Informative InDel Markers in a Mapping Population of Pusa 1103 × ILWC 46

The comparison of NGS genome resequencing of high (Pusa 1103 and HPNB) and low (ILWC 46 and LPNB) pod number mapping parental accessions and bulks with reference *kabuli* (CDC Frontier) genomic sequence discovered 25,477 and 24,166 high-quality InDel markers (Tables S1, S2). This included 8628 InDel markers exhibiting polymorphism between high (Pusa 1103 and HPNB) and low (ILWC 46 and LPNB) pod number mapping parents and bulks according to their congruent physical positions (bp) on the reference *kabuli* genome (Figures 1A, 2A, Table S3). Notably, *in-silico* fragment length polymorphism detected by markers based on their size (bp) of InDels varied from 1 to 18 bp with a mean of 3.1 bp. More than 73.1% (6306) InDel markers exhibited 1–3 bp *in-silico* fragment length polymorphism while remaining 26.9% (2322) markers revealed 4–18 bp fragment length polymorphism (Table S3).

The 5965 and 2663 markers of the total designed 8628 InDel markers were physically mapped on eight chromosomes and unanchored scaffolds of *kabuli* chickpea genome with an average map density of 63.1 kb [varying from 37.7 (chromosome 4) to 94.4 (chromosome 6) kb] (Figures 1A, 2A, Table S3). Highest and lowest number of InDel markers were mapped on chromosomes 4 (1304 markers with a mean map density: 37.7 kb) and 8 (293 markers with a mean map density: 56.2 kb), respectively (Figure 2A, Table S3). The structural annotation of 8628 InDel markers revealed the occurrence of 6642 (77%) and 1986 (23%) markers in the intergenic regions and different sequence components of 1523 protein-coding genes, respectively (Figures 1B, 2A, Table S3). The average frequency of InDel markers within genes was estimated as 1.3 markers/gene. A maximum of 1124 (56.6%) gene-derived InDel markers were designed from the DRRs (downstream regulatory regions) of 945 genes and minimum of 30 (1.5%) markers derived from the URRs (upstream regulatory regions) of 23 genes (Figures 1C, 2A, Table S3). Remarkably, 33 and 35 coding InDel markers developed from the 33 and 34 genes caused frameshift mutations and affected initiation/stop codons (large-effect mutations), respectively.

The KOG-based functional annotation of 1523 genes with 1986 InDel markers exhibited primary roles of 1095

(55.1%) markers-carrying 815 genes in multiple cellular, biological, and molecular processes in crop plants (Table S3). This revealed enrichment of InDel markers-containing genes basically involved in post-translational modification, protein turnover, and chaperones (O, 111 markers in 81 genes, 10.1%), transcription (K, 74 markers in 49 genes, 6.7%), and signal transduction mechanisms (T, 70 markers in 55 genes, 6.4%), beside general function prediction (R, 200 markers in 158 genes, 18.3%; Table S3). Of the 799 (40.2%) InDel markers developed from 603 TF-encoding genes (representing 50 TF gene families), the genes belonging to MYB (90 markers in 62 genes, 11.3%), bHLH (78 markers in 61 genes, 9.8%), C2H2 zinc finger (51 markers in 36 genes, 6.4%), and NAC (48 markers in 36 genes, 6%) TF families were abundant (Table S3).

Development of Genome-Wide Informative InDel Markers in a Mapping Population of Pusa 256 × ILWC 46

We developed 15,640 and 17,077 high-quality InDel markers by comparing the NGS genome resequencing of high (Pusa 256 and HPNB) and low (ILWC 46 and LPNB) pod number mapping parental accessions and bulks with reference *kabuli* (CDC Frontier) genomic sequence (Tables S4, S5). This included 5263 InDel markers revealing polymorphism between high (Pusa 256 and HPNB) and low (ILWC 46 and LPNB) pod number mapping parents and bulks according to their congruent physical positions (bp) on the reference *kabuli* genome (Figures 1A, 2B, Table S6). Notably, *in-silico* fragment length polymorphism detected by markers based on their size (bp) of InDels varied from 1 to 18 bp with a mean of 3.1 bp. More than 73% (3842) InDel markers showed 1–3 bp *in-silico* fragment length polymorphism while remaining 27% (1421) markers revealed 4–18 bp fragment length polymorphism (Table S6). The 3402 and 1861 markers of the total designed 5263 InDel markers were physically mapped on eight chromosomes and unanchored scaffolds of *kabuli* chickpea genome with an average map density of 103.5 kb [ranging from 61.9 (chromosome 4) to 148.7 (chromosome 5) kb] (Figures 1A, 2B, Table S6). Highest and lowest number of InDel markers were mapped on chromosomes 4 (795 markers with a mean map density: 61.9 kb) and 8 (218 markers with a mean map density: 75.6 kb), respectively. The structural annotation of 5263 InDel markers revealed the presence of 4168 (79.2%) and 1095 (20.8%) markers in the intergenic regions and different sequence components of 868 protein-coding genes, respectively (Figures 1B, 2B, Table S6). The mean frequency of InDel markers within genes was measured as 1.3 markers/gene. A maximum of 576 (52.6%) gene-derived InDel markers were designed from the DRRs of 487 genes and minimum of 24 (2.2%) markers derived from the URRs of 19 genes (Figures 1C, 2B, Table S6). Remarkably, 24 and 26 coding InDel markers developed from the 24 and 26 genes caused frameshift mutations and affected initiation/stop codons (large-effect mutations), respectively. A total of 13,891 including 8628 and 5263 polymorphic InDel markers identified from two inter-specific mapping populations of Pusa 1103

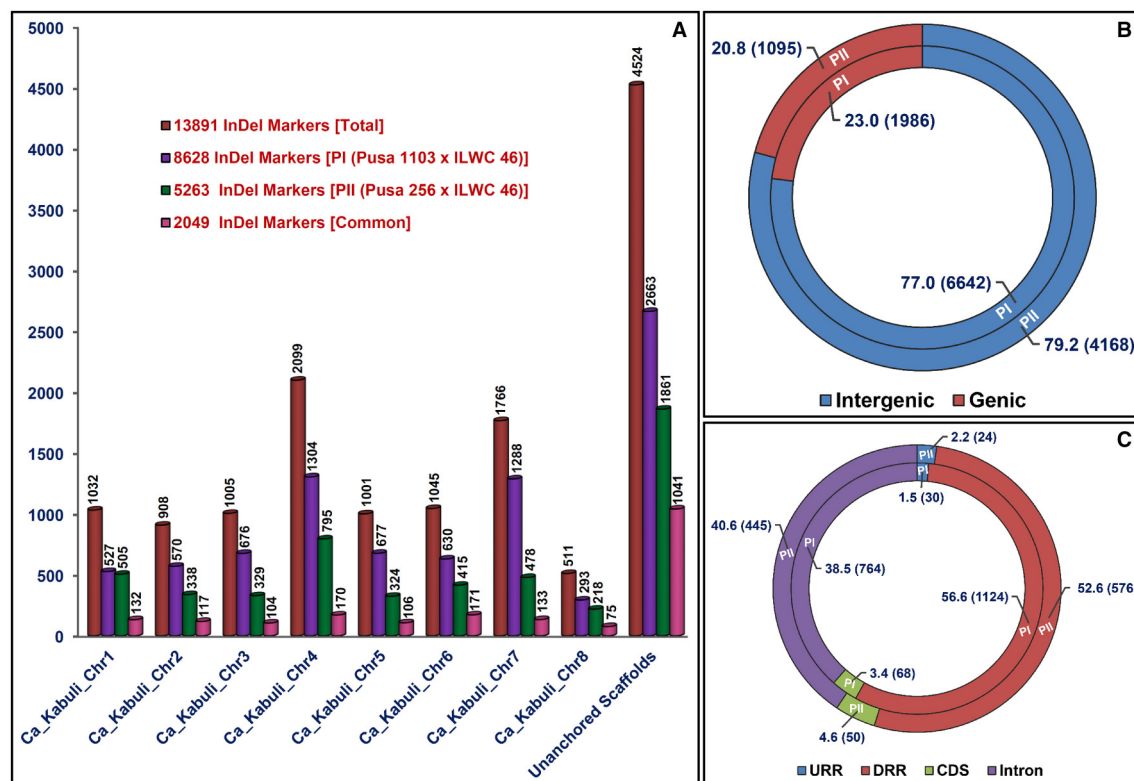


FIGURE 1 | Genetic constitution and genome-wide distribution pattern of InDel markers exhibiting differentiation between high and low pod number parental accessions and homozygous bulks from each of two mapping populations [PI: (Pusa 1103 × ILWC 46) and PII: (Pusa 256 × ILWC 46)] with respect to *kabuli* chickpea (CDC Frontier) genome. (A) These InDel markers were physically mapped on eight chromosomes and unanchored scaffolds of *kabuli* chickpea genome, which are illustrated by bar diagrams. (B,C) Relative frequency (proportionate distribution) of InDel markers designed from the intergenic as well as diverse coding (CDS) and non-coding (introns, URRs, and DRRs) sequence components of genes annotated from *kabuli* chickpea genome. Parenthesis designates the number of InDel markers developed from each sequence regions of *kabuli* genome. The CDS (coding DNA sequences), URR (upstream regulatory region), and DRR (downstream regulatory region) of genes were defined as per the gene annotation of *kabuli* chickpea genome (v).

× ILWC 46 and Pusa 256 × ILWC 46, respectively were compared/correlated. This included 2049 markers common between these two mapping populations based on congruent marker physical positions on the *kabuli* genome (Figure 1A, Tables S3, S6, S7).

The KOG-based functional annotation of 868 genes with 1095 InDel markers exhibited primary roles of 597 (54.5%) markers-carrying 466 genes in multiple cellular, biological, and molecular processes in crop plants (Table S6). This revealed enrichment of InDel markers-containing genes basically involved in post-translational modification, protein turnover, and chaperones (O, 59 markers in 46 genes, 9.9%) and signal transduction mechanisms (T, 37 markers in 30 genes, 6.2%), beside general function prediction (R, 87 markers in 75 genes, 14.6%) and unknown function (S, 35 markers in 29 genes, 5.9%; Table S6). Of the 440 (40.2%) InDel markers developed from 337 TF-encoding genes (representing 44 TF gene families), the genes belonging to *MYB* (55 markers in 36 genes, 12.5%), *NAC* (40 markers in 31 genes, 9.1%), *bHLH* (39 markers in 31 genes, 8.9%), C2H2 zinc finger (29 markers in 17 genes, 6.6%), and *B3* (24 markers in 21 genes, 5.4%) TF families were predominant (Table S6).

Experimental Validation of InDel Markers to Access Their Amplification and Polymorphic Potential among Cultivated and Wild Chickpea Accessions

To access the amplification and polymorphic potential of designed InDel markers, 3743 markers exhibiting ≥4 bp *in-silico* fragment length polymorphism between the parental accessions and bulks of two inter-specific mapping populations, were selected for experimental validation using the gel- and PCR amplicons resequencing-based assays. These markers were PCR amplified and genotyped using the genomic DNA of three mapping parental chickpea accessions (Pusa 1103, Pusa 256, and ILWC 46), from which the InDel markers were originally discovered. Notably, 3612 of 3743 markers produced single reproducible PCR amplicons in agarose gel with a mean amplification success rate of 96.5% (Figure 3). Of these, 3413 (94.5%) amplified markers revealing *in-silico* polymorphism at least between two combination of mapping parental chickpea accessions were got validated experimentally using both agarose gel- and PCR amplicons resequencing-based assays. The PCR amplicons resequencing-led validation and genotyping of InDel

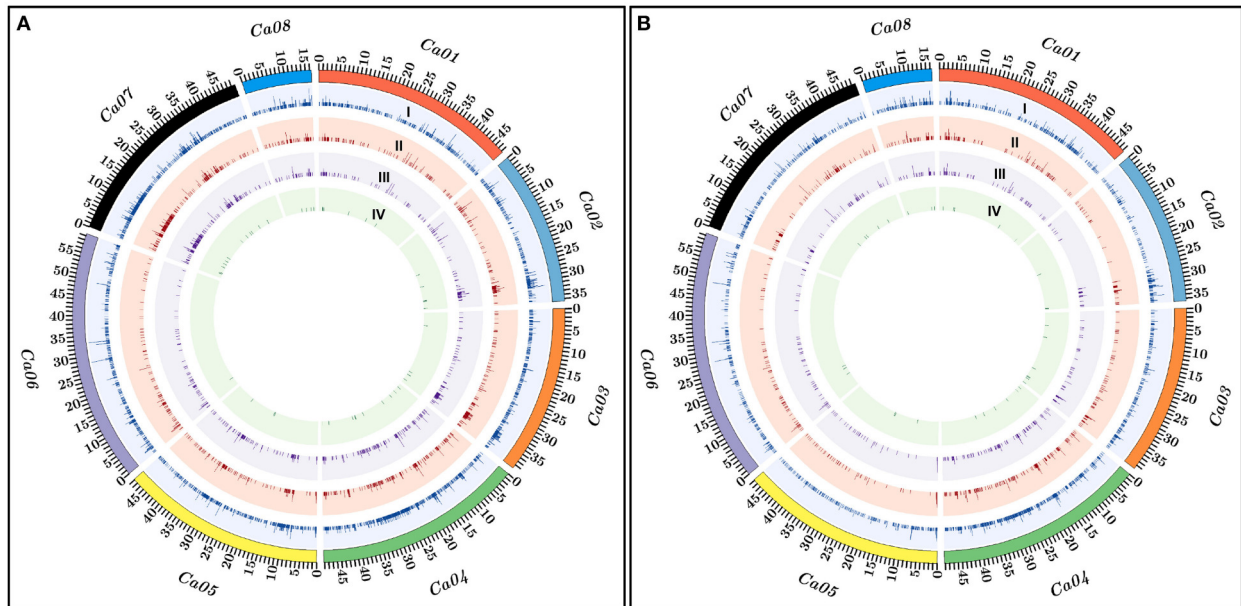


FIGURE 2 | The relative genomic distribution of PI (Pusa 1103 × ILWC 46) (A) and PII (Pusa 256 × ILWC 46) (B) mapping populations-derived polymorphic InDel markers physically mapped on eight chromosomes of *kabuli* chickpea genome are depicted by the Circos circular ideograms. The outermost circles represent the different physical sizes (Mb) of eight chromosomes coded with multiple colors as per the pseudomolecule sizes documented in *kabuli* chickpea genome (Varshney et al., 2013). Total InDel markers (I) including gene-derived (II), regulatory (III), and coding (IV) markers polymorphic between high and low pod number parental accessions and homozygous bulks of two inter-specific mapping populations—PI (Pusa 1103 × ILWC 46) (A) and PII (Pusa 256 × ILWC 46) (B) are indicated.

markers ascertained the presence of expected InDels, which further corresponded well with their *in-silico* fragment length polymorphism detected among three mapping parental chickpea accessions.

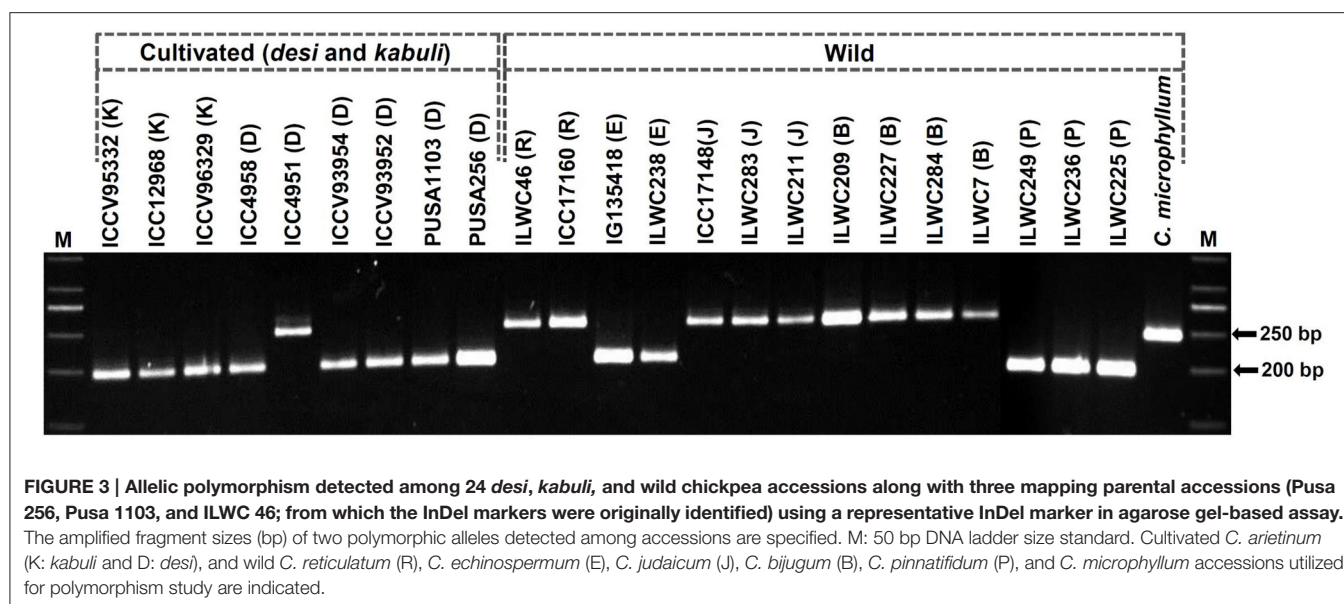
To evaluate the potential of InDel markers for detecting polymorphism among accessions, large-scale genotyping of 3413 polymorphic InDel markers were performed in a diverse set of 24 *desi*, *kabuli* and wild chickpea accessions (Figure 3). These markers overall generated 6849 alleles among accessions with an average PIC of 0.71. Three thousand three hundred-two (96.7%, mean PIC: 0.65) of 3413 markers were found to be polymorphic among cultivated and wild chickpea accessions, whereas 2831 (83%, mean PIC: 0.60) markers exhibited polymorphism among cultivated *desi* and *kabuli* accessions. Interestingly, 2355 (69%) markers exhibited polymorphism among six *desi* accessions (1–2 alleles with a mean PIC: 0.57), whereas 1980 (58%) markers revealed polymorphism among three *kabuli* accessions (1–2 alleles with a mean PIC: 0.51). A set of 2969 (87%) InDel markers exhibited polymorphism among 15 accessions belonging to six annual/perennial wild species—*Cicer reticulatum*, *C. echinospermum*, *C. judaicum*, *C. bijugum*, *C. pinnatifidum*, and *C. microphyllum* of primary, secondary, and tertiary gene-pools.

Generation of a Consensus High-Density Inter-Specific Chickpea Genetic Linkage Map

For constructing high-resolution inter-specific genetic linkage maps, 1059 and 594 InDel markers revealing polymorphism between high and low pod number parental accessions (Pusa

1103 vs. ILWC 46 and Pusa 256 vs. ILWC 46) and bulks (HPNB vs. LPNB) were genotyped among 102 and 98 individuals of two F₅ mapping populations—PI (Pusa 1103 × ILWC 46) and PII (Pusa 256 × ILWC 46), respectively. The linkage analysis using these marker genotyping data mapped 1059 and 594 InDel markers across eight LGs of two PI and PII mapping populations-derived inter-specific chickpea genetic maps, respectively (Figure 4, Table 1). In a PI mapping population-derived genetic map, highest and lowest numbers of InDel markers were mapped on CaLG07 (235 markers) and CaLG08 (47 markers), respectively (Table 1). In another PII mapping population-derived genetic map, the CaLG04 (128 markers) and CaLG08 (31 markers) contained maximum and minimum number of mapped InDel makers, respectively (Table 1). The eight LGs-based two inter-specific genetic maps generated from PI and PII mapping populations spanned total map-lengths of 978.21 and 603.26 cM, with the mean inter-marker distances of 0.92 and 1.01 cM, respectively (Table 1). Longest map-lengths spanning 221.34 and 122.28 cM were observed in CaLG07 and CaLG04 of PI and PII mapping populations-derived genetic linkage maps, respectively. The CaLG01 and CaLG04 of genetic linkage maps constructed from PI and PII mapping populations had most saturated genetic maps with the mean inter-marker distances of 0.81 and 0.99 cM, respectively (Table 1).

Combining the genotyping information of 1653 including 1059 and 594 InDel markers mapped genetically on two PI and PII mapping populations-derived inter-specific genetic linkage maps, respectively, we constructed a consensus high-resolution



genetic linkage map of chickpea. A set of 174 InDel markers found common between two inter-specific genetic linkage maps were served as the anchor markers for integration and defining the linkage groups/chromosomes of these genetic maps. A consensus 1479 InDel markers-anchored inter-specific genetic linkage map was generated, which covered a total map-length of 978.61 cM with an average inter-marker distance (map-density) of 0.66 cM (**Figure 4, Table 1**). The map-density of a consensus inter-specific genetic map varied from 0.50 cM (CaLG01) to 0.81 cM (CaLG05). Highest (297) and lowest (67) number of InDel markers were mapped on CaLG07 and CaLG08 spanning longest and shortest map-lengths of 221.34 and 41.59 cM, respectively (**Table 1**).

Molecular Mapping of Major Pod Number and Seed Yield QTLs

For molecular mapping of pod number and seed yield per plant QTLs, primarily the genetic inheritance pattern of PN and SYP traits in two inter-specific mapping populations was determined. A significant difference of PN (5–237) with 13–14.8% CV and 80–81% H^2 in 102 and 98 individuals and parental accessions of two inter-specific F_5 mapping populations of Pusa 1103 (PN: 129) \times ILWC 46 (29) and Pusa 256 (125) \times ILWC 46 (29) was observed. Moreover, the parental accessions and individuals belonging to these two mapping populations of Pusa 1103 (SYP: 38.3 g) \times ILWC 46 (19.2 g) and Pusa 256 (37.4 g) \times ILWC 46 (19.2 g) exhibited a significant difference of SYP (16.7–54.3 g) with 9.7–10.3% CV and 78–80% H^2 . The continuous variation as well as normal frequency distribution along with bi-directional transgressive segregation of PN and SYP were observed in these both mapping populations reflecting the quantitative genetic inheritance pattern of traits under study. A significant positive correlation between PN and SYP traits based on Pearson's correlation coefficient ($r = 69$ –72%) was evident.

Two years multi-location replicated field phenotyping data of PN and genotyping information of 1059 and 594 InDel markers genetically mapped on eight chickpea chromosomes of two inter-specific genetic linkage maps constructed from PI (Pusa 1103 \times ILWC 46) and PII (Pusa 256 \times ILWC 46) mapping populations, respectively were integrated for molecular mapping of major PN QTLs. This analysis detected three major genomic regions harboring three robust QTLs associated with PN trait, which were mapped on chromosomes 2 and 4 of each PI and PII mapping populations-derived inter-specific genetic maps (**Figure 4, Table 2**). For PI mapping population-derived high-resolution genetic linkage map, three major genomic regions underlying three PN QTLs (*Caq^aPN2.1*, *Caq^aPN4.1*, and *Caq^aPN4.2*) spanned (7.55–8.99 cM on chromosome 4) with 51 InDel markers, were mapped on chromosomes 2 and 4 (**Figure 4, Table 2**). The individual major PN QTL explained 18–25% phenotypic variation (R^2) for pod number trait at an 8.5–12.7 LOD. The PVE (phenotypic variation explained) measured for all three major PN QTLs in combination was 38.4%. All three major PN QTLs exhibited positive additive gene effect (2.7–4.3) of pod number trait with major allelic contribution from a high pod number parental chickpea accession Pusa 1103. For PII mapping population-derived high-density genetic linkage map, three major genomic regions underlying three PN QTLs (*Caq^bPN2.1*, *Caq^bPN4.1*, and *Caq^bPN4.2*) spanned (5.45–7.71 cM on chromosome 4) with 33 InDel markers, were mapped on chromosomes 2 and 4 (**Figure 4, Table 2**). The individual major PN QTL explained 15–22% phenotypic variation (R^2) for pod number trait at a 6.7–11.4 LOD. The PVE measured for all three major PN QTLs in combination was 34.1%. All three major PN QTLs exhibited positive additive gene effect (3.3–4.7) of pod number trait with major allelic contribution from a high pod number parental chickpea accession Pusa 256. Further, a high-density consensus genetic linkage map constructed by integrating two inter-specific genetic linkage

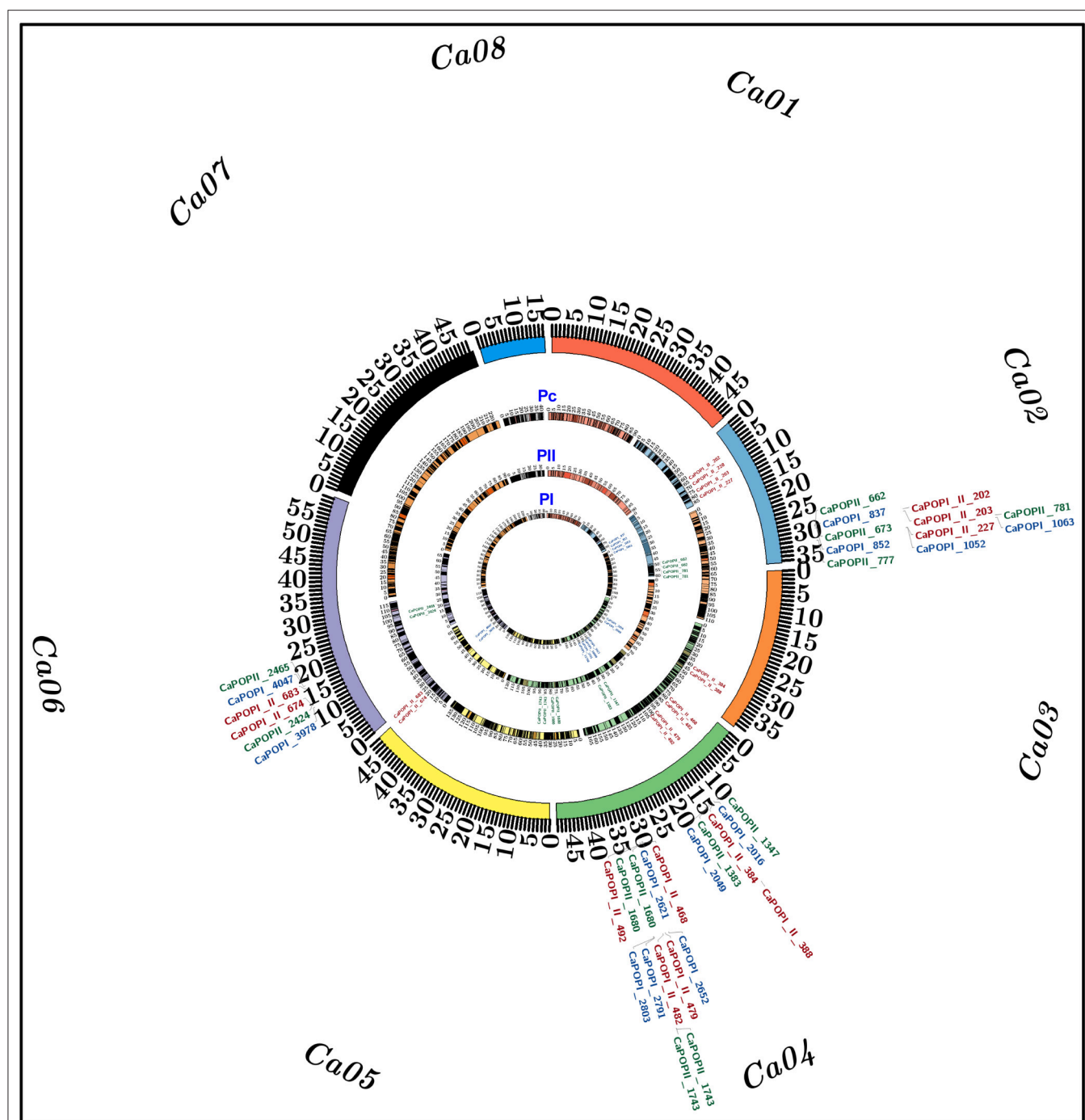


FIGURE 4 | The identified three of each major PN and SYP QTLs mapped on chromosomes 2, 4, and 6 of two high-density 1059 and 594 InDel markers-anchored inter-specific genetic linkage maps (PI: Pusa 1103 × ILWC 46) and (PII: Pusa 256 × ILWC 46) and a consensus 1479 InDel markers-led high-resolution genetic map (Pc) of chickpea, are illustrated by the Circos circular ideograms (PI, PII, and Pc). The circles represent the different genetic map length (cM) (spanning 5–10 cM uniform genetic distance intervals between bins) of eight LGs/chromosomes coded with multiple colors. The integration of a consensus genetic map (Pc) with physical map at the identified three of each major PN and SYP QTLs scaled-down the long genomic regions harboring these major QTLs into short PN and SYP robust QTL physical intervals (indicated with red color InDel markers) mapped on *kabuli* chickpea chromosomes 2, 4, and 6. The InDel markers flanking the six major PN and SYP QTLs mapped on chromosomes 2, 4, and 6 of high-resolution genetic maps—PI, PII, and Pc are highlighted with blue, green, and red color, respectively. The detail information on QTLs and InDel markers are provided in the **Table 2**. The outermost circles denote the various physical sizes (Mb) of eight chromosomes coded with multiple colors as per the pseudomolecule sizes documented in *kabuli* chickpea genome (Varshney et al., 2013).

TABLE 1 | InDel markers mapped on eight chromosomes of two inter-specific genetic linkage maps [(Pusa 1103 × ILWC 46) and (Pusa 256 × ILWC 46)] and a consensus inter-specific chickpea genetic linkage map of chickpea.

Linkage groups (LGs)/ chromosomes (Chr)	InDel markers mapped			Map length covered (cM)			Map-density [mean inter-marker distance (cM)]		
	(Pusa 1103 × ILWC 46)	(Pusa 256 × ILWC 46)	A consensus genetic linkage map	(Pusa 1103 × ILWC 46)	(Pusa 256 × ILWC 46)	A consensus genetic linkage map	(Pusa 1103 × ILWC 46)	(Pusa 256 × ILWC 46)	A consensus genetic linkage map
LG(Chr)01	111	91	180	89.94	90.34	90.34	0.81	0.99	0.50
LG(Chr)02	88	59	127	87.79	60.74	87.79	1.0	1.03	0.69
LG(Chr)03	128	67	178	110.98	70.74	110.98	0.87	1.05	0.62
LG(Chr)04	185	128	288	169.91	122.28	169.91	0.92	0.95	0.59
LG(Chr)05	134	62	170	138.71	68.07	138.71	1.03	1.10	0.81
LG(Chr)06	132	67	172	117.95	67.61	117.95	0.90	1.01	0.68
LG(Chr)07	235	89	297	221.34	90.35	221.34	0.94	1.01	0.74
LG(Chr)08	47	31	67	41.59	33.13	41.59	0.88	1.07	0.62
Total	1059	594	1479	978.21	603.26	978.61	0.92	1.01	0.66

maps was utilized as an anchor for molecular mapping of major PN QTLs in chickpea. This identified three major genomic regions underlying three PN QTLs (*Caq^cPN2.1*, *Caq^cPN4.1*, and *Caq^cPN4.2*) spanned (2.7–5.7 cM on chromosome 4) with 33 InDel markers, which were mapped on three different genomic regions on chromosomes 2 and 4 (Figure 4, Table 2). The individual major PN QTL explained 20–28% phenotypic variation (R^2) for pod number trait at a 9.4–13.8 LOD. The PVE measured for all three major PN QTLs in combination was 39.7%. All three major PN QTLs exhibited positive additive gene effect (3.8–4.5) of pod number trait with major allelic contribution from the high pod number parental chickpea accessions Pusa 1103/Pusa 256.

The integration of two individual and consensus inter-specific genetic maps with that of physical maps of *kabuli* genome exhibited the common occurrences of five and six InDel markers at the identified three major PN QTL regions of chromosomes 2 and 4, respectively, among these genetic maps based on congruent physical positions on the *kabuli* genome (Figure 4, Table 2). These consensus three major PN QTL regions spanning short physical intervals (*Caq^cPN2.1*: 29,445,927–32,393,633 bp, *Caq^cPN4.1*: 13,509,527–14,397,300 bp, and *Caq^cPN4.2*: 31,806,633–33,714,267 bp) were got validated in two diverse inter-specific mapping populations. These were thus considered as promising major candidate genomic regions underlying robust QTLs governing pod number to be deployed for marker-assisted genetic enhancement of chickpea (Figure 4, Table 2, Tables S1–S6). The structural and functional annotation of these delineated three major short PN QTL intervals of 2.94 (*Caq^cPN2.1*), 0.89 (*Caq^cPN4.1*), and 1.91 (*Caq^cPN4.2*) Mb exhibited the presence of a total 27, 5, and 14 InDel markers including 19, 4, and 13 markers in the intergenic regions and 8, 1, and 1 markers in the different sequence components of 7, 1, 1 genes annotated from *kabuli* chickpea genome, respectively (Tables S1–S4). At these three major PN robust QTL intervals, especially nine regulatory InDel markers-containing genes corresponding to diverse transcription factors (TFs; like DUF1677, LBD, WRKY, and C2H2 zinc finger) and multiple

cellular metabolism-related proteins such as cytochrome P450 and ubiquitin were identified, which can possibly serve as candidates for quantitative dissection of complex pod number trait in chickpea (Table S8).

In order to evaluate the efficacy of identified major PN QTLs in governing seed yield, multi-location/years replicated field phenotyping data of SYP trait were integrated with genotyping information of InDel markers genetically mapped on chromosomes of PI (Pusa 1103 × ILWC 46) and PII (Pusa 256 × ILWC 46) mapping populations-derived two inter-specific genetic linkage maps for molecular mapping of major SYP QTLs in chickpea. This identified three major genomic regions underlying three robust QTLs associated with SYP trait which were mapped on chromosomes 2, 4, and 6 of each PI and PII mapping populations-led inter-specific genetic maps (Figure 4, Table 2). For PI mapping population-based high-resolution genetic linkage map, three major genomic regions underlying three robust SYP QTLs (*Caq^aSYP2.1*, *Caq^aSYP4.1*, and *Caq^aSYP6.1*) covered (7.3 cM on chromosome 2–10.5 cM on chromosome 4) with 57 InDel markers were detected (Figure 4, Table 2). The individual major SYP QTL explained 16–23% phenotypic variation (R^2) for seed yield trait at an 8.0–12.3 LOD. The PVE estimated for all three major SYP QTLs in combination was 30%. All of these identified three SYP QTLs exhibited positive additive gene effect (2.5–4.1) of seed yield trait with major allelic contribution from a high SYP parental chickpea accession Pusa 1103. For PII mapping population-led high-density genetic map, three major genomic regions harboring three robust YP QTLs (*Caq^bSYP2.1*, *Caq^bSYP4.1*, and *Caq^bSYP6.1*) spanned (4.9 cM on chromosome 2–7.7 cM on chromosome 4) with 38 InDel markers were identified (Figure 4, Table 2). The individual major SYP QTL explained 17–23% phenotypic variation (R^2) for yield trait at a 7.5–10.6 LOD. The PVE measured for all three major SYP QTLs in combination was 32%. All of these three major SYP QTLs exhibited positive additive gene effect (3.1–4.0) of seed yield trait with major allelic contribution from a high SYP parental chickpea accession Pusa 256. The use of a high-density consensus genetic linkage map (constructed by integrating two

TABLE 2 | Molecular mapping of major pod number and seed yield per plant QTLs in chickpea.

InDel markers-anchored genetic linkage maps	*QTLs	LGs/Chromosomes	Marker intervals with genetic positions (cM)	QTL physical intervals (bp)	LOD	PVE (%)	A
PI; (Pusa 1103 × ILWC 46)	<i>Caq²PN2.1</i>	CaLG(Chr)2	CaPOPII_837 (73.96)–CaPOPII_1052 (81.99)	CaPOPII_837 (29,220,253)–CaPOPII_1052 (32,393,633)	8.5	18	2.7
	# <i>Caq²PN4.1</i>	CaLG(Chr)4	CaPOPII_2016 (45.19)–CaPOPII_2049 (52.74)	CaPOPII_2016 (13,068,436)–CaPOPII_2049 (15,252,487)	12.7	25	3.9
	# <i>Caq²PN4.2</i>	CaLG(Chr)4	CaPOPII_2652 (108.09)–CaPOPII_2803 (117.08)	CaPOPII_2652 (31,259,145)–CaPOPII_2803 (33,857,426)	11.5	21	4.3
	<i>Caq²SYP2.1</i>	CaLG(Chr)2	CaPOPII_852 (75.50)–CaPOPII_1063 (82.80)	CaPOPII_852 (29,829,940)–CaPOPII_1063 (32,713,895)	8.0	16	2.5
	<i>Caq²SYP4.1</i>	CaLG(Chr)4	CaPOPII_2621 (106.11)–CaPOPII_2791 (116.58)	CaPOPII_2621 (30,685,617)–CaPOPII_2791 (33,714,297)	11.8	23	3.0
	<i>Caq²SYP6.1</i>	CaLG(Chr)6	CaPOPII_3978 (24.42)–CaPOPII_4047 (34.05)	CaPOPII_3978 (12,253,200)–CaPOPII_4047 (17,086,650)	12.3	20	4.1
PII; (Pusa 256 × ILWC 46)	<i>Caq²PN2.1</i>	CaLG(Chr)2	CaPOPII_673 (49.17)–CaPOPII_777 (54.07)	CaPOPII_673 (29,445,927)–CaPOPII_777 (32,379,585)	6.7	15	3.3
	# <i>Caq²PN4.1</i>	CaLG(Chr)4	CaPOPII_1347 (30.69)–CaPOPII_1383 (36.14)	CaPOPII_1347 (12,225,286)–CaPOPII_1383 (14,397,300)	11.4	22	4.7
	# <i>Caq²PN4.2</i>	CaLG(Chr)4	CaPOPII_1680 (78.06)–CaPOPII_1743 (85.77)	CaPOPII_1680 (31,101,573)–CaPOPII_1743 (34,169,705)	9.8	19	4.2
	<i>Caq²SYP2.1</i>	CaLG(Chr)2	CaPOPII_662 (48.55)–CaPOPII_781 (54.10)	CaPOPII_662 (29,072,040)–CaPOPII_781 (32,393,633)	7.5	17	3.1
	<i>Caq²SYP4.1</i>	CaLG(Chr)4	CaPOPII_1680 (78.06)–CaPOPII_1743 (85.77)	CaPOPII_1680 (31,101,573)–CaPOPII_1743 (34,169,705)	10.6	23	4.0
	<i>Caq²SYP6.1</i>	CaLG(Chr)6	CaPOPII_2424 (14.81)–CaPOPII_2465 (21.08)	CaPOPII_2424 (12,877,302)–CaPOPII_2465 (18,331,022)	9.5	20	3.8
Pc; [PI (Pusa 1103 × ILWC 46) and PII (Pusa 256 × ILWC 46)]	<i>Caq²PN2.1</i>	CaLG(Chr)2	CaPOPII_IL_202 (61.8)–CaPOPII_IL_228 (68.3)	CaPOPII_IL_202 (29,445,927)–CaPOPII_IL_228 (32,393,633)	9.4	20	3.8
	# <i>Caq²PN4.1</i>	CaLG(Chr)4	CaPOPII_IL_384 (40.3)–CaPOPII_IL_388 (43.0)	CaPOPII_IL_384 (13,509,527)–CaPOPII_IL_388 (14,397,300)	13.8	28	4.5
	# <i>Caq²PN4.2</i>	CaLG(Chr)4	CaPOPII_IL_479 (94.9)–CaPOPII_IL_492 (100.6)	CaPOPII_IL_479 (31,806,633)–CaPOPII_IL_492 (33,714,267)	11.7	24	4.0
	<i>Caq²SYP2.1</i>	CaLG(Chr)2	CaPOPII_IL_203 (75.50)–CaPOPII_IL_227 (81.95)	CaPOPII_IL_203 (29,829,940)–CaPOPII_IL_227 (32,379,585)	10.2	22	4.1
	<i>Caq²SYP4.1</i>	CaLG(Chr)4	CaPOPII_IL_468 (76.64)–CaPOPII_IL_482 (80.89)	CaPOPII_IL_468 (30,535,832)–CaPOPII_IL_482 (32,227,293)	14.5	27	4.3
	<i>Caq²SYP6.1</i>	CaLG(Chr)6	CaPOPII_IL_674 (20.2)–CaPOPII_IL_683 (26.0)	CaPOPII_IL_674 (12,877,302)–CaPOPII_IL_683 (16,547,931)	12.4	25	4.0

**Caq²PN2.1* (C. arietinum PI-derived pod number QTL on chromosome 2 number 1), *Caq²PN2.1* (C. arietinum PI-derived pod number QTL on chromosome 2 number 1), *Caq²PN2.1* (C. arietinum PI-derived seed yield per plant QTL on chromosome 2 number 1), *Caq²PN2.1* (C. arietinum PI-derived seed yield per plant QTL on chromosome 2 number 1) and *Caq²PN2.1* (C. arietinum PI-derived seed yield per plant QTL on chromosome 2 number 1). PVE, Proportion of phenotypic variation explained by QTLs, A, additive effect; positive additive effect inters alleles from high PN and SYP mapping parental chickpea accessions (Pusa 1103 and Pusa 256). Details regarding InDel markers are mentioned in the Tables S1–S8. *documented previously by Das et al. (2016).

inter-specific genetic linkage maps) as an anchor identified three major genomic regions underlying three robust SYP QTLs (*Caq^cSYP2.1*, *Caq^cSYP4.1*, and *Caq^cSYP6.1*) covered (4.2 cM on chromosome 4–6.4 cM on chromosome 2) with 50 InDel markers which were mapped on three different genomic regions on chromosomes 2, 4, and 6 (**Figure 4, Table 2**). The individual major SYP QTL explained 22–27% phenotypic variation (R^2) for seed yield trait at a 10.2–14.5 LOD. The PVE estimated for all three major SYP QTLs in combination was 35%. All of these three major SYP QTLs exhibited positive additive gene effect (4.0–4.3) of seed yield trait with major allelic contribution from the high SYP parental chickpea accessions Pusa 1103/Pusa 256.

We integrated two individual and consensus inter-specific genetic maps with that of physical maps of *kabuli* genome which exhibited common occurrence of 25, 15, and 10 InDel markers at the identified three major SYP robust QTL regions of chromosomes 2, 4, and 6, respectively, among these genetic maps based on congruent physical positions on the *kabuli* genome (**Figure 4, Table 2**). These consensus three major SYP QTL regions spanning short physical intervals (*Caq^cSYP2.1*: 29,829,940–32,379,585 bp, *Caq^cSYP4.1*: 30,535,832–32,227,293 bp, and *Caq^cSPN6.1*: 12,877,302–16,547,931 bp) were got validated in two diverse inter-specific mapping populations. Therefore, we considered these QTL intervals as promising major candidate genomic regions underlying robust QTLs regulating seed yield trait which could be deployed for marker-assisted genetic enhancement of chickpea (**Figure 4, Table 2, Tables S1–S6**). The structural and functional annotation of these delineated three major short SYP QTL intervals of 2.55 (*Caq^cSYP2.1*), 1.69 (*Caq^cSYP4.1*), and 3.67 (*Caq^cSYP6.1*) Mb revealed the presence of 25, 15, and 10 InDel markers including 17, 11, and 9 markers in the intergenic regions and 8, 4, and 1 markers in the different sequence components of genes annotated from *kabuli* chickpea genome, respectively (Tables S1–S4). At these three major SYP robust QTL regions, especially three regulatory and one coding nonsense non-synonymous InDel markers-containing genes corresponding to diverse TFs such as *bHLH*, *LBD*, *WRKY*, and *NAC* were identified. These molecular tags can be considered as candidates for dissection of complex seed yield quantitative trait in chickpea (Table S8).

DISCUSSION

The pod number is a complex yield component quantitative trait, which is known to be regulated by multiple genes/QTLs in chickpea (Kujur et al., 2015a,b; Das et al., 2016). For more efficient dissection of this complex trait, the present study essentially utilized two inter-specific mapping populations [(Pusa 1103 × ILWC 46) and (Pusa 256 × ILWC 46)] with contrasting PN trait to construct a high-density InDel markers-anchored consensus inter-specific genetic linkage map for molecular mapping of major PN QTLs in chickpea. To attain these major objectives, high-throughput whole genome NGS resequencing data with a high mean *kabuli* genome (64.1%, 474 Mb) and sequencing depth (~11.6-fold) coverage, generated from high and low pod number parental accessions as well as bulks of two inter-specific mapping populations, were normalized/compared

to develop high-quality accurate InDel markers at a genome-wide scale. The reliability of these identified InDels was ascertained by their potential to differentiate both high and low pod number mapping parental accessions as well as bulks constituted from two studied inter-specific mapping populations. This implicates the robustness of strategy developed in our study for mining and developing valid non-erroneous InDel markers at a genome-wide scale by comparing the resequences among parents and bulks of mapping populations. Consequently, 82,360 markers targeting these non-spurious informative InDels, discriminating the *desi* (Pusa 1103 and Pusa 256), *kabuli* (CDC Frontier) and wild (ILWC 46) accessions from each other including 13,891 markers differentiating the high and low PN—mapping parental accessions (Pusa 1103 vs. ILWC 46 and Pusa 256 vs. ILWC 46) and bulks, were developed in chickpea. These informative InDel markers were structurally and functionally annotated in diverse sequence components of genome/genes (TFs), thereby can be deployed for manifold genomics-assisted breeding applications in chickpea. The observed InDel marker-based genetic polymorphism expectedly infers close evolutionary relatedness of *desi* rather than *kabuli* with wild chickpea (Abbo et al., 2005; Berger et al., 2005; Toker, 2009; Jain et al., 2013; Varshney et al., 2013; Saxena et al., 2014a; Bajaj et al., 2015b; Das et al., 2015; Kujur et al., 2015a,b). We detected almost an identical range (1–18 bp) and mean (3.1 bp) level of InDel-based *in-silico* fragment length polymorphism between two inter-specific mapping populations [(Pusa 1103 × ILWC 46) and (Pusa 256 × ILWC 46)] developed by using a common (wild *C. reticulatum* accession: ILWC 46) as well as two different (*desi* accessions: Pusa 1103 and Pusa 256) parental accessions of chickpea. This is possibly because of common ancestry between Pusa 1103 and Pusa 256, since Pusa 1103 has been developed from the multiple inter-cross between wild *C. reticulatum* and *desi* chickpea accessions involving Pusa 256 as one of the parent (Bharadwaj et al., 2011). Therefore, close progenitor relatedness and similar genetic backgrounds of these two different parental accessions that were used to develop mapping populations could have influenced the detection of identical InDel polymorphism level (bp) between two studied inter-specific mapping populations of chickpea. The cost-efficient user-friendly InDel markers especially developed from the regulatory and coding (frameshift/large-effect mutations) regions of genes/TFs possibly have a greater impact on transcriptional gene regulation (expression) resulting alteration of gene functions in chickpea. These functionally relevant InDel markers thus have a broader practical application in establishing efficient marker-trait linkages and quick identification of potential genes/QTLs governing vital agronomic traits in chickpea.

The inter (97%)- and intra (58–87%)-specific polymorphic potential detected by the InDel markers among *desi*, *kabuli*, and wild chickpea accessions is much higher/comparable to that estimated especially with the sequence-based SSR, SNP, and InDel markers (Nayak et al., 2010; Bharadwaj et al., 2011; Gujaria et al., 2011; Agarwal et al., 2012; Hiremath et al., 2012; Kujur et al., 2013, 2015a,b,c; Deokar et al., 2014; Saxena et al., 2014a,b; Bajaj et al., 2015a,b,c; Das et al., 2015). The informative genome-wide

InDel markers, especially those resolved/genotyped by a simpler cost-effective agarose gel-based assay, exhibiting high intra-specific polymorphic potential among accessions belonging to cultivated and wild chickpea are highly significant. These informative markers could thus serve as a beneficial genomic resource for their immense use in high-throughput genetic analysis including marker-assisted introgression breeding and genetic enhancement of chickpea.

We generated two high-resolution (mean inter-marker distances: 0.92 and 1.01 cM) 1059 and 594 InDel markers-anchored inter-specific genetic linkage maps [(Pusa 1103 × ILWC 46) and (Pusa 256 × ILWC 46)] and a high-resolution (0.66 cM) 1479 InDel markers-led consensus genetic map of chickpea. The map-densities estimated for these genetic maps are comparable with that documented so far in multiple intra- and inter-specific mapping populations-derived genetic maps of chickpea (Nayak et al., 2010; Gujaria et al., 2011; Hiremath et al., 2012; Kujur et al., 2013, 2015c; Deokar et al., 2014; Saxena et al., 2014b; Bajaj et al., 2015a; Das et al., 2015). Therefore, the genetic linkage maps constructed by us have potential to identify and map major QTLs governing various stress tolerance and yield component traits including pod number in chickpea.

The two inter-specific mapping populations utilized in the present study for major PN and SYP QTL mapping revealed a wider phenotypic variability and higher heritability (consistent phenotypic expression) across geographical locations/years for pod number and seed yield per plant trait. These mapping populations thus can serve as a useful genetic resource for molecular mapping of major PN and SYP QTLs in chickpea. The quantitative genetic inheritance pattern of PN and SYP trait was evident from its continuous variation and transgressive segregation as well as normal frequency distribution in the two studied inter-specific mapping populations. This infers the involvement of multiple genes/QTLs in controlling PN and SYP trait in these two diverse mapping populations of chickpea. The identification of strong trait-associated robust QTLs that are well-validated in multiple genetic backgrounds (inter-specific mapping populations in the present study) is essential for efficient deployment of informative markers tightly linked to these QTLs in marker-assisted genetic enhancement of chickpea. To detect well-validated robust PN and SYP QTLs, three major genomic regions underlying each of PN and SYP QTLs mapped individually on two inter-specific genetic maps derived from two mapping populations [(Pusa 1103 × ILWC 46) and (Pusa 256 × ILWC 46)] were compared and correlated. This led to identification of three redundant major genomic regions with short physical intervals of 2.94 (*Caq^cPN2.1*), 0.89 (*Caq^cPN4.1*), and 1.91 (*Caq^cPN4.2*) Mb for PN as well as 2.55 (*Caq^cSYP2.1*), 1.69 (*Caq^cSYP4.1*), and 3.67 Mb (*Caq^cSYP6.1*) for SYP mapped on chromosomes 2, 4, and 6 of a high-density consensus inter-specific genetic linkage map. The validation of three major PN and SYP QTLs across two diverse inter-specific mapping populations indicated their significance as robust QTLs to be utilized in marker-assisted selection and genetic enhancement of chickpea. Three of each long major PN (2.17–3.17 Mb) and SYP (2.88–5.45 Mb) QTL intervals mapped on two individual inter-specific genetic linkage maps were scaled-down into three short major genomic regions underlying robust PN (0.89–2.94

Mb) and SYP (1.69–3.67 Mb) QTLs on a high-density consensus genetic map. This implicates the potential utility of a high-density consensus inter-specific genetic linkage map for high-resolution molecular mapping/fine mapping of robust QTLs and delineation of potential candidate genes governing PN and SYP trait in chickpea.

To ascertain the novelty of three detected robust PN QTLs, the major genomic regions underlying these QTLs were compared with that reported in previous QTL mapping studies employing diverse inter- and intra-specific chickpea mapping populations. Based on congruent physical positions on chickpea chromosomes, two robust PN QTLs (*Caq^cPN4.1* and *Caq^cPN4.2*) exhibited correspondence with two known major PN QTLs (*Caq^cPN4.1* and *Caq^cPN4.2*) identified earlier from similar two inter-specific mapping populations using the mQTL-seq strategy in chickpea (Das et al., 2016). The remaining one robust PN QTL (*Caq^cPN2.1*) identified by us has not been reported so far by any QTL mapping studies and thus considered as a novel QTL regulating pod number in chickpea. This could be due to use of genome-wide InDel markers in the present investigation for traditional QTL mapping vis-à-vis whole genome SNP markers for mQTL-seq analysis in the past study (Das et al., 2016). Notably, two robust PN QTLs (*Caq^cPN4.1* and *Caq^cPN4.2*) and one novel PN QTL (*Caq^cPN2.1*) spanning 0.89–2.94 Mb physical intervals, mapped on the chromosomes 2 and 4 of a high-density inter-specific consensus genetic map, were targeted by us to delineate potential candidate genes regulating pod number in chickpea. Interestingly, one identified novel major PN robust QTL (*Caq^cPN2.1*) revealing correspondence with a major SYP QTL (*Caq^cSYP2.1*), was colocalized based on their congruent physical positions on chickpea chromosome 2. The structural and functional annotation of these short physical PN and SYP QTL regions with available *kabuli* genome sequence especially identified 12 informative InDel markers-led regulatory and one nonsense non-synonymous natural allelic variants in multiple candidate genes/TFs which are known to be the key players of growth and development in diverse crop plants (Moon et al., 2004; Libault et al., 2009; Agarwal et al., 2011; Bartel, 2012; Sadanandom et al., 2012; Bajaj et al., 2015c; Xu et al., 2015). These functionally relevant molecular tags possibly regulating pod number and seed yield delineated by us can be deployed in marker-assisted genetic enhancement to develop high seed and pod-yielding cultivars with increased pod/seed number and yield in chickpea.

Summarily, the current investigation was able to provide multiple novel outcomes vis-à-vis our previous study (Das et al., 2016) that utilizes similar two inter-specific mapping populations for pod number QTL mapping in chickpea. Our study optimized a strategy to develop high-quality informative InDel markers especially from the low coverage NGS genome resequencing data of multiple mapping parental accessions and homozygous bulks/individuals at a genome-wide scale in chickpea with limited resource expenses. In addition, more than three thousand InDel markers exhibiting high intra-/inter-specific polymorphic potential among cultivated (*desi* and *kabuli*) and wild accessions even by a simpler economical agarose gel-based assay were screened for their effective utilization in genomics-assisted breeding applications of chickpea with narrow

genetic base. The efficacy of InDel markers for construction of a high-density inter-specific consensus genetic linkage map and molecular mapping of high-resolution major pod number robust QTLs was demonstrated. Despite using similar mapping populations, we scanned two alike and an additional novel major QTL governing pod number between our present and past studies in chickpea. The detection of novel major pod number QTL is possibly due to deployment of high-resolution InDel markers-based traditional QTL mapping in the current study by genotyping of genome-wide InDel markers individually among segregating lines of two mapping populations of chickpea. Efforts have been made to establish efficient correlation between our identified major pod number and seed yield robust QTLs to ascertain the efficacy of these PN QTLs in marker-assisted genetic enhancement for developing high seed and pod-yielding chickpea cultivars. Essentially, the user-friendly InDel markers tightly linked to the genes underlying three (two previously reported and one novel in this study) major pod number as well as three novel seed yield robust QTLs delineated by us can be utilized in marker-assisted foreground selection for efficient screening of numerous back-cross mapping individuals especially by a cost-effective agarose gel-based assay in order to complement the ongoing chickpea molecular breeding program. Among these, especially the InDel markers developed from the genes colocalized at both novel pod number and seed yield robust QTL regions exhibiting increased major allelic effect for combined high PN and SYP traits appear much promising to be utilized in chickpea

genomics-assisted breeding. This will eventually drive marker-aided genetic enhancement to develop chickpea cultivars with high seed and pod number and yield in laboratories with minimal infrastructural facilities.

AUTHOR CONTRIBUTIONS

RS conducted all experiments, bioinformatics analysis and drafted the manuscript. MS and DB involved in InDel markers genotyping and allelic diversity data analysis. MS developed mapping populations and helped in their multilocation phenotyping. SP conceived and designed the study, guided data analysis and interpretation, participated in drafting and correcting the manuscript critically and gave the final approval of the version to be published. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by the core grant of National Institute of Plant Genome Research (NIPGR), New Delhi, India.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.01362>

REFERENCES

- Abbo, S., Gopher, A., Rubin, B., and Lev-Yadun, S. (2005). On the origin of near Eastern founder crops and the 'dump-heap hypothesis'. *Genet. Res. Crop Evol.* 52, 491–495. doi: 10.1007/s10722-004-7069-x
- Agarwal, G., Jhanwar, S., Priya, P., Singh, V. K., Saxena, M. S., Parida, S. K., et al. (2012). Comparative analysis of *kabuli* chickpea transcriptome with *desi* and wild chickpea provides a rich resource for development of functional markers. *PLoS ONE* 7:e52443. doi: 10.1371/journal.pone.0052443
- Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33, 189–202. doi: 10.1002/bies.201000107
- Bajaj, D., Das, S., Badoni, S., Kumar, V., Singh, M., Bansal, K. C., et al. (2015b). Genome-wide high-throughput SNP discovery and genotyping for understanding natural (functional) allelic diversity and domestication patterns in wild chickpea. *Sci. Rep.* 5:12468. doi: 10.1038/srep12468
- Bajaj, D., Saxena, M. S., Kujur, A., Das, S., Badoni, S., Tripathi, S., et al. (2015c). Genome-wide conserved non-coding microsatellite (CNMS) marker-based integrative genetical genomics for quantitative dissection of seed weight in chickpea. *J. Exp. Bot.* 66, 1271–1290. doi: 10.1093/jxb/eru478
- Bajaj, D., Upadhyaya, H. D., Khan, Y., Das, S., Badoni, S., Shree, T., et al. (2015a). A combinatorial approach of comprehensive QTL-based comparative genome mapping and transcript profiling identified a seed weight-regulating candidate gene in chickpea. *Sci. Rep.* 5:9264. doi: 10.1038/srep09264
- Bartel, B. (2012). Focus on ubiquitin in plant biology. *Plant Physiol.* 160:1. doi: 10.1104/pp.112.203208
- Berger, J. D., Buck, R., Henzell, J. M., and Turner, N. C. (2005). Evolution in the genus *Cicer* vernalisation response and low temperature pod set in chickpea (*C. arietinum* L.) and its annual wild relatives. *Aust. J. Agric. Res.* 56, 1191–1200. doi: 10.1071/AR05089
- Bharadwaj, C., Srivastava, R., Chauhan, S. K., Satyavathi, C. T., Kumar, J., Faruqui, A., et al. (2011). Molecular diversity and phylogeny in geographical collection of chickpea (*Cicer* sp.) accessions. *J. Genet.* 90, e94–e100. doi: 10.1007/s12041-011-0114-6
- Bohra, A., Saxena, R. K., Gnanesh, B. N., Saxena, K., Byregowda, M., Rathore, A., et al. (2012). An intra-specific consensus genetic map of pigeonpea [*Cajanus cajan* (L.) Millspaugh] derived from six mapping populations. *Theor. Appl. Genet.* 125, 1325–1338. doi: 10.1007/s00122-012-1916-5
- Das, S., Singh, M., Srivastava, R., Bajaj, D., Saxena, M. S., Rana, J. C., et al. (2016). mQTL-seq delineates functionally relevant candidate gene harbouring a major QTL regulating pod number in chickpea. *DNA Res.* 23, 53–65. doi: 10.1093/dnares/dsv036
- Das, S., Upadhyaya, H. D., Srivastava, R., Bajaj, D., Gowda, C. L. L., Sharma, S., et al. (2015). Genome-wide insertion-deletion (InDel) marker discovery and genotyping for genomics-assisted breeding applications in chickpea. *DNA Res.* 22, 377–386. doi: 10.1093/dnares/dsv020
- Deokar, A. A., Ramsay, L., Sharpe, A. G., Diapari, M., Sindhu, A., Bett, K., et al. (2014). Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC Genomics* 15:708. doi: 10.1186/1471-2164-15-708
- Gujaria, N., Kumar, A., Dauthal, P., Dubey, A., Hiremath, P., Bhanu Prakash, A., et al. (2011). Development and use of genic molecular markers (GMMs) for construction of a transcript map of chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 122, 1577–1589. doi: 10.1007/s00122-011-1556-1
- Hiremath, P. J., Kumar, A., Penmetsa, R. V., Farmer, A., Schlueter, J. A., Chamarthi, S. K., et al. (2012). Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnol. J.* 10, 716–732. doi: 10.1111/j.1467-7652.2012.00710.x

- Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tpj.12173
- Kudapa, K., Azam, S., Sharpe, A. G., Taran, B., Li, R., Deonovic, B., et al. (2014). Comprehensive transcriptome assembly of chickpea (*Cicer arietinum* L.) using Sanger and next generation sequencing platforms: development and applications. *PLoS ONE* 9:e86039. doi: 10.1371/journal.pone.0086039
- Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–374. doi: 10.1093/dnares/dst015
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015a). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Front. Plant Sci.* 6:162. doi: 10.3389/fpls.2015.00162
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015b). A genome-wide SNP scan accelerates trait-regulatory genomic loci identification in chickpea. *Sci. Rep.* 5:11166. doi: 10.1038/srep11166
- Kujur, A., Upadhyaya, H. D., Shree, T., Bajaj, D., Das, S., Saxena, M., et al. (2015c). Ultra-high density intra-specific genetic linkage maps accelerate identification of functionally relevant molecular tags governing important agronomic traits in chickpea. *Sci. Rep.* 5:9468. doi: 10.1038/srep09468
- Libault, M., Joshi, T., Benedito, V. A., Xu, D., Udvardi, M. K., and Stacey, G. (2009). Legume transcription factor genes: what makes legumes so special? *Plant Physiol.* 151, 991–1001. doi: 10.1104/pp.109.144105
- Li, Y. H., Liu, B., Reif, J. C., Liu, Y. L., Li, H. H., Chang, R. Z., et al. (2014). Development of insertion and deletion markers based on biparental resequencing for fine mapping seed weight in soybean. *Plant Genome* 7, 1–8. doi: 10.3835/plantgenome2014.04.0014
- Moghaddam, M. S., Song, Q., Mamidi, S., Schmutz J., Lee R., Cregan P., et al. (2014). Developing market class specific InDel markers from next generation sequence data in *Phaseolus vulgaris* L. *Front. Plant Sci.* 5:185. doi: 10.3389/fpls.2014.00185
- Moon, J., Parry, G., and Estelle, M. (2004). The ubiquitin-proteasome pathway and plant development. *Plant Cell* 16, 3181–3195. doi: 10.1105/tpc.104.161220
- Nayak, S. N., Zhu, H., Varghese, N., Datta, S., Choi, H. K., Horres, R., et al. (2010). Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome. *Theor. Appl. Genet.* 120, 1415–1441. doi: 10.1007/s00122-010-1265-1
- Sadanandom, A., Bailey, M., Ewan, R., Lee, J., and Nelis, S. (2012). The ubiquitin-proteasome system: central modifier of plant signalling. *N. Phytol.* 196, 13–28. doi: 10.1111/j.1469-8137.2012.04266.x
- Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014b). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031
- Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014a). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484
- Token, C. (2009). A note on the evolution of *kabuli* chickpeas as shown by induced mutations in *Cicer reticulatum* Ladizinsky. *Genet. Resour. Crop. Evol.* 56, 7–12. doi: 10.1007/s10722-008-9336-8
- Van Ooijen, J. W. (2009). *MapQTL 6: Software for the Mapping of Quantitative Trait Loci in Experimental Populations of Diploid Species*. Wageningen: Kyazma, B. V.
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Varshney, R. K., Thudi, M., Nayak, S. N., Gaur, P. M., Kashiwagi, J., Krishnamurthy, L., et al. (2014). Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* 127, 445–462. doi: 10.1007/s00122-013-2230-6
- Wang, Y., Lu, J., Chen, S., Shu, L., Palmer, R. G., Xing, G., et al. (2014). Exploration of presence/absence variation and corresponding polymorphic markers in soybean genome. *J. Integr. Plant Biol.* 56, 1009–1019. doi: 10.1111/jipb.12208
- Xu, J., Wang, X. Y., and Guo, W. Z. (2015). The cytochrome P450 superfamily: key players in plant development and defense. *J. Int. Agri.* 14, 1673–1686. doi: 10.1016/S2095-3119(14)60980-1

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Srivastava, Singh, Bajaj and Parida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Selection of Novel Cowpea Genotypes Derived through Gamma Irradiation

Lydia N. Horn^{1,2}, Habteab M. Ghebrehiwot^{1,3*} and Hussein A. Shimelis^{1,3}

¹ School of Agricultural, Earth and Environmental Sciences, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa, ² Directorate of Research and Training, Plant Production Research, Ministry of Agriculture, Water and Forestry, Windhoek, Namibia, ³ African Centre for Crop Improvement, University of KwaZulu-Natal, Pietermaritzburg, South Africa

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University
of Mexico, Mexico

Reviewed by:

Karthika Rajendran,
International Center for Agricultural
Research in the Dry Areas, Morocco
Minviluz Garcia Stacey,
University of Missouri, USA

*Correspondence:

Habteab M. Ghebrehiwot
ghebrehiwot@ukzn.ac.za

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 23 November 2015

Accepted: 19 February 2016

Published: 10 March 2016

Citation:

Horn LN, Ghebrehiwot HM
and Shimelis HA (2016) Selection
of Novel Cowpea Genotypes Derived
through Gamma Irradiation.
Front. Plant Sci. 7:262.
doi: 10.3389/fpls.2016.00262

Cowpea (*Vigna unguiculata* [L.] Walp.) yields are considerably low in Namibia due to lack of improved varieties and biotic and abiotic stresses, notably, recurrent drought. Thus, genetic improvement in cowpea aims to develop cultivars with improved grain yield and tolerance to abiotic and biotic stress factors. The objective of this study was to identify agronomically desirable cowpea genotypes after mutagenesis using gamma irradiation. Seeds of three traditional cowpea varieties widely grown in Namibia including Nakare (IT81D-985), Shindimba (IT89KD-245-1), and Bira (IT87D-453-2) were gamma irradiated with varied doses and desirable mutants were selected from M₂ through M₆ generations. Substantial genetic variability was detected among cowpea genotypes after mutagenesis across generations including in flowering ability, maturity, flower and seed colors and grain yields. Ten phenotypically and agronomically stable novel mutants were isolated at the M₆ each from the genetic background of the above three varieties. The selected promising mutants' lines are recommended for adaptability and stability tests across representative agro-ecologies for large-scale production or breeding in Namibia or similar environments. The novel cowpea genotypes selected through the study are valuable genetic resources for genetic enhancement and breeding.

Keywords: cowpea, gamma radiation, mutation breeding, mutants, legume improvement

INTRODUCTION

Cowpea (*Vigna unguiculata* L. Walp.) is a leguminous species used as food, forage, and vegetable crop mainly in the tropics (Steele, 1972). The grains are an excellent source of food and feed; a vital nutrient for healthy growth both for humans and livestock. The leaves, green pods, and grains are consumed as a dietary source of protein (Ghaly and Alkoik, 2010). The cowpea grain contains 23% protein and 57% carbohydrate, and the leaves contain 27–34% of proteins. The crop originated and domesticated in Southern Africa, which was later spread to east and West Africa and Asia (International Institute for Tropical Agriculture [IITA], 2004). In semi-arid West and Central Africa, it is consumed as a pulse where it supplements the daily diet (Bressani, 1985). Thus, cowpea production remains the most prominent food legume cultivated by farmers majorly in most sub-Saharan African countries. The main reasons being the natural ability of the crop to withstand moderate episodes of drought and its adaptation to grow in nutrient limited soils. Cowpea is also able to fix atmospheric nitrogen in marginal soils where farmers are unable to adequately fertilize

their crops due to unaffordability or inaccessibility (Steele, 1972). Accounts indicate that greater than 16,000 genotypes of cowpea are registered in trust for the World Bank by the International Institute of Tropical Agriculture, (IITA) Ibadan, Nigeria. Such a huge genotype bank is believed to provide a wide range of information on the agronomy and potential benefits of the crop.

The southern African region is reportedly considered the centre of diversity of *V. unguiculata* which includes Namibia, Botswana, Zambia, Zimbabwe, Mozambique, and the Republic of South Africa (Ng and Marachel, 1985). In Namibia, cowpea is the second most important crop next to pearl millet. Nearly, 95% of the smallholder farmers in the northern part of the country grow cowpea for food security and/or livelihoods. However, cowpea yields of the available cultivars are considerably low (250–350 kg/ha) predominantly due to lack of improved varieties and biotic and abiotic stresses notably recurrent severe drought. Hence, genetic improvement in cowpea requires systematic breeding and development of genotypes associated with higher yielding capacity and drought resilience.

Genetic variation is the basis for plant breeding programs. Most conventional crop improvement programs rely on natural genetic variation present among germplasm pools (Ceccarelli and Grando, 2007). Mutations can be induced in various ways, such as exposure of plant propagules, including seeds, tissues, and organs, to physical and chemical mutagens (Mba et al., 2010). Induced mutagenesis has the potential to create genetic variation for genetic enhancement and breeding in a relatively shorter time unlike natural mutation or controlled crosses of especially unrelated parents (Singh et al., 2006; Wani, 2006; Tulmann Neto et al., 2011). Gnanamurthy et al. (2012) reported that induced mutations have been successfully used in breeding of seed propagated crops since 1940s. The Mutant Varieties Database (MVD) of FAO (Food and Agriculture Organisation of the United Nations) and the International Atomic Energy Agency (IAEA) maintained a list of 2,252 crop cultivars developed through artificial mutations (Nielen, 2004). These cultivars were released across 59 countries worldwide, mainly in the continental Asia (1,142 cultivars), Europe (847), and North America (160) (Maluszynski, 2001; Maluszynski et al., 2009). Studies indicate that induced mutagenesis has successfully modified several plant traits such as plant height, maturity, seed shattering resistance, disease resistance, oil quality and quantity, malting quality, size and quality of starch granules of cowpea (Goyal and Khan, 2010; Singh et al., 2013).

In South Africa, cowpea mutants were developed through selections from the M_2 to M_4 generations. These included the drought tolerant mutants such as 447, 217, and 346, and mutants such as 447, MA2, and 217 isolated for their high yielding ability under well-watered conditions (De Ronde and Spreeth, 2007). Furthermore, early maturing cowpea mutants with leaflets containing tendrils, broad leaves, and light green pods were developed through gamma irradiation in Nigeria (Adekola and Oluleye, 2007). The use of gamma irradiation at different doses has been reported to change the proximate and anti-nutritive compositions in pulses (Udensi et al., 2012). Some varieties of groundnut were developed in Congo through gamma irradiation (Tshilenge-Lukanda et al., 2012). Wani (2006)

reported a significant increase in the mean values of the fertile branches per plant, pods per plant and seed yield per plant (SYP) in mutant varieties of mungbean (*Vigna radiata* [L.] Wilczek) derived through gamma irradiation.

In light of this, a collaborative research was developed in 2009 between the Namibian Government and the IAEA under Technical Cooperation project on induced mutation breeding using Gamma irradiation. This created a platform for pre-breeding and breeding of high yielding, drought tolerant and insect pest resistant genotypes of cowpea. Gamma irradiation was recommended by the Namibian Radiation Regulatory Authority as an alternative option to create new crop genotypes in a short period of time without any negative impact to the environment.

Therefore, the objective of this study was to identify desirable cowpea genotypes after gamma irradiation of three traditional cowpea varieties widely grown in Namibia including Nakare (IT81D-985), Shindimba (IT89KD-245-1), and Bira (IT87D-453-2) through continuous selections from M_2 through M_6 generations.

MATERIALS AND METHODS

Plant Material and Gamma Irradiation

Three cowpea genotypes widely grown in Namibia, namely, Nakare (IT81D-985), Shindimba (IT89KD-245-1) and Bira (IT87D-453-2) were obtained from Likorerere Farmers Co-operatives at Kavango Region, Namibia. The seeds were irradiated at the International Atomic Energy Agency (IAEA), Agriculture and Biotechnology Laboratory, A-2444 Seibersdorf, Austria using a CO60 source Gammacell Model No. 220. Various doses were used to establish the optimum irradiation level that can achieve optimum mutation frequency with least possible and unintended damage. The three varieties were gamma irradiated as follows: Bira [0, 75, 150, 300, 450, and 600 Gy], Nakare [0, 100, 150, 200, 250, and 300 Gy] and Shindimba [0, 100, 150, 200, 300, and 400 Gy]. Preliminary tests showed that the three varieties differed in their optimal requirement of irradiation doses and was used as the bases for using different doses for each genotype (Horn and Shimelis, 2013). The 0 Gy dose served as a comparative control.

Study Sites, Experimental Design, and Field Establishment

A series of selection experiments were conducted at three different sites; namely Mannheim, Bagani, and Omahenene. Mannheim Research Station is located in Oshikoto region along the north central of Namibia and it is situated at an altitude of 1234 m above sea level (masl). Bagani Research Station is located at (1007 masl) north east in the Kavango East region, whereas Omahenene research station is situated in the Omusati Region in North-Western Namibia at altitude of 1109 masl. In general, climatic, biological conditions of the selection sites vary considerably. Physicochemical properties of the sites are provided in **Table 1**. The M_1 and M_2 generations were evaluated at Mannheim Research Station during the 2009/2010 and 2010/2011 seasons, respectively. The M_3 generations were established at

TABLE 1 | Physicochemical properties of soils at Mannheim, Bagani, and Omahenene research sites in Namibia.

Sample/parameter	Research Station (study site)		
	Mannheim	Bagani	Omahenene
Soil pH	7.87	7.5	8.2
Total Nitrogen%	0.06	0.06	0.05
Organic carbon%	0.38	0.48	0.60
Phosphorus (ppm)	18	58.2	14
Potassium me%	0.17	0.9	0.99
Calcium me %	1.6	1.3	1.38
Magnesium me%	4.74	1.7	4.80
Manganese me%	0.05	0.18	0.17
Copper (ppm)	0.6	0.6	0.5
Iron (ppm)	0.5	0.7	0.5
Zinc (ppm)	0.6	0.5	0.4
Sodium %	0.10	0.09	0.07
EC mS/cm	0.29	0.18	0.36

ppm, part per million; me, milliequivalent; EC, electrical conductivity.

Bagani research station during the 2011/2012 season. The M₄ and M₅ were established at Omahenene Research Station in 2012/2013 and 2013/2014 season, respectively.

Plots were arranged in a randomized complete block design using two replications. Plants were established using intra-row spacing of 20 cm and inter-row spacing of 75 cm. Seedlings were thinned to one plant per hill after 2 weeks from planting. Weeds were controlled manually. Planting of the M₁ seeds was done under normal growing conditions with supplemental irrigation during dry spell. Each row of the M₁ generation contained 26 individuals, making a total of 104 plants per irradiation dose. At harvest the M₂ seeds were bulked in separate bags according to irradiation doses (Figure 1). During the M₂ to M₅ generations' variable number of individual plants ranging from 50 to 100 per irradiation dose were assayed for qualitative and quantitative observations.

Selection Procedure and Data Collection

The selection procedure was undertaken based on methods adapted from Maluszynski et al. (2009). The selection procedure used in the study is illustrated in Figure 1. The irradiated seeds (M₁) were planted in the field at Mannheim research station under standard cultural practices. All the pods, from the M₁ plants that survived were harvested and bulked according to their respective radiation doses and genotypes. Consequently, the harvested M₂ seeds were planted in the field at Mannheim as M₂ population during 2010/2011 season in the form of progeny rows for individual plant selection and to develop the M₃ seeds. The M₃ seed from selected M₂ plants were planted at Omahenene and Bagani Research Station during 2011/2012 for evaluation. The M₃ plants at both sites were evaluated in the field using morphological and agronomical attributes. Pods from selected M₃ plants were harvested. During 2012/2013, the M₄ seeds obtained from the selected M₃ population were planted at Omahenene Research Station as single-plant progenies and segregants were selected with desired traits. During 2013/2014

the M₅ seeds obtained from the selected M₄ population were planted at Omahenene Research Station as single-plant progenies and selection were made toward desired trait on single plant basis. Uniform, non-segregating mutant progenies, were bulked at this stage to hasten the breeding cycle. During 2014/2015 the M₆ generation was evaluated at Omahenene, Bagani, and Mannheim using suitable lines selected for seed yield and related traits.

Data Collection and Analysis

Both quantitative and qualitative data were collected during evaluations from the M₂ to M₅ generations. The data collected included: days to 50% germination (DG), percent seed emergence (ES%), number of abnormal individuals or visual phenotype mutants (ABN), total number of surviving plants per plot (TNP), number of main branches (NMB) averaged over 10 randomly selected and tagged plants, days to 50% flowering (DTF), days to 50% pod setting (DPS), days to 50% maturity (DMT), number of pods per plant (NPP) averaged over five pods per selected plant, pod length (PL) expressed in cm and averaged over five pods per plant, pod weight per plant (PW) in gram, number of seeds per pod (NSP) averaged over five pods per plant, 100 seed weight (HSW) in gram and SYP in gram. The qualitative data collected included variation in flower color (FC) and seed color (SC) during the M₁ and M₂ generations. Additional qualitative data such as, pod shape (PS), pod color (PC), seed coat texture (SCT), and growth habit (GH) were collected from M₂ to M₅ generations. Data were analyzed and descriptive statistics summarized using the SAS statistical program (SAS, 2002).

RESULTS

Phenotypic Characterization of Mutants

Qualitative and Quantitative Traits at M₁ and M₂

During the M₁ and M₂ generations the percentage field establishment (ES) ranged between 79 to 89%, respectively (Table 2). Nakare and Shindimba mutants had ES of 0% at irradiation doses of 250, 300, and 400 Gy. Phenotypic abnormalities such as albinism, leaf deformity, single stem, seedless pods or short pod sizes were invariably observed at the following doses and genotypes: 450 and 600 Gy (Bira); 150 and 200 Gy (Nakare); and 100, 150, and 200 Gy (Shindimba) (Figure 2). Segregation of FC (white and purple) were observed at the M₂ with the following doses and genotypes: 300, 450, and 600 Gy (Bira), 100 and 200 Gy (Nakare), and 100, 150, and 200 Gy (Shindimba) (Figure 3). SC variations were observed during the M₂ (Figure 4). White, brown, red, and cream SC were common in Bira mutants across all irradiation doses. In addition to these Nakare and Shindimba had speckled, chocolate, light brown, black, mixed and dark brown SC when subjected to irradiation doses of 100, 150, and 200 Gy (Table 2 and Figure 4). Bira mutants displayed relatively high seed yields varying from 98 to 200 g/plant at 0 and 600 Gy, respectively (Table 2).

Qualitative Traits Evaluated during the M₃ to the M₅

Variable number of individual plants was available for selection during M₃ to M₅ generations, because of the strength of

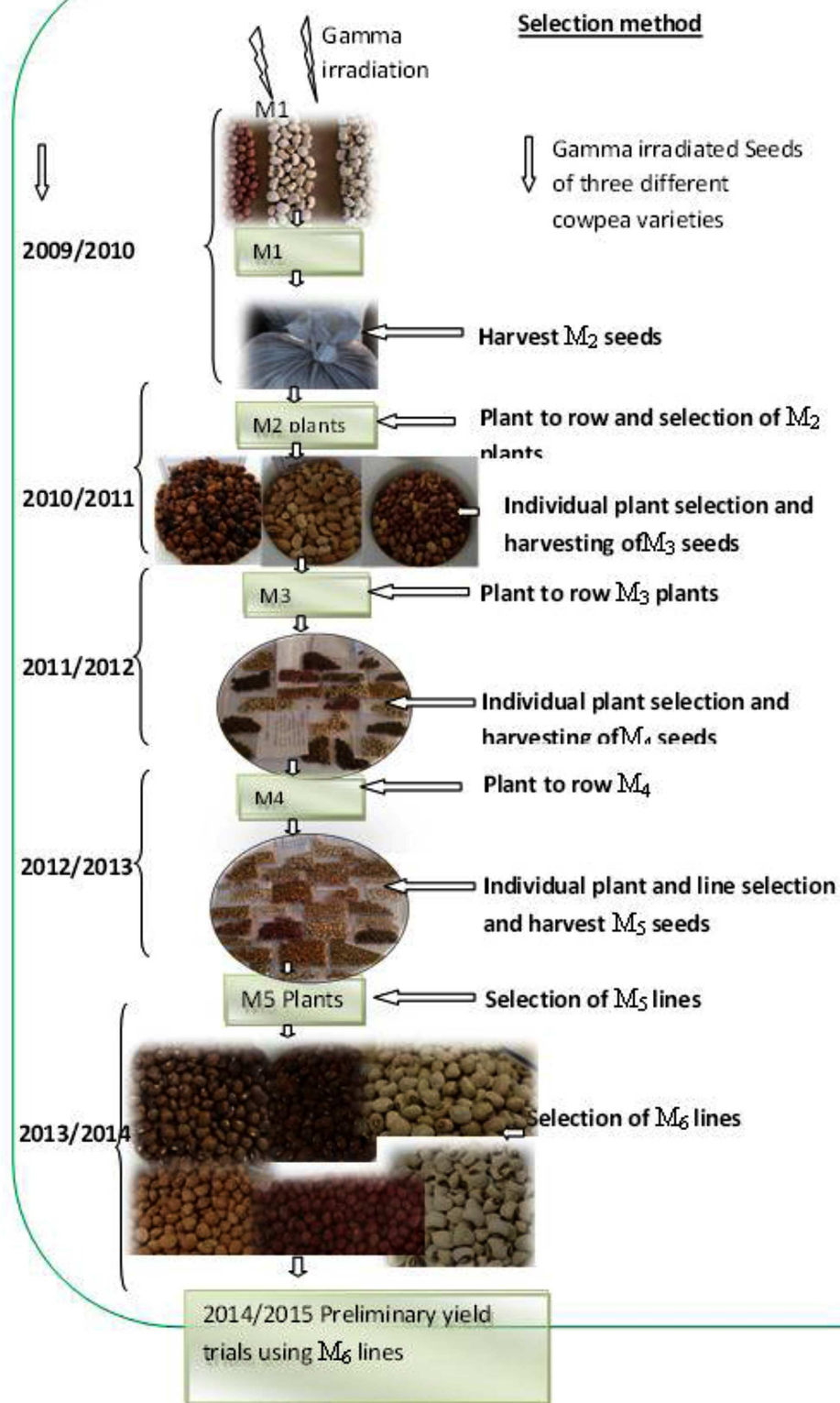


FIGURE 1 | Illustration of the selection methods during 2009–2014.

TABLE 2 | Phenotypic characteristics of mutants observed during the first two seasons 2009/2010 and 2010/2011 at Mannheim Research Station.

Variety	M ₁ (2009/2010)						M ₂ (210/2011)				
	Dose (Gy)	ES%	ABN	FC	SC	SYP	ES%	ABN	FC	SC	SYP
Bira	0	89	0	2	3	2.9	99	0	2	3	98
	75	80	0	2	3	2.9	88	0	2	1,2,3,4	150
	150	87	0	2	3	3.1	89	0	2	1,2,3,4	162
	300	82	1	2	3	2.0	90	1,2,3	1,2	1,2,3,4	160
	450	81	1,2	2	3	1.6	93	1,2,3	1,2	1,2,3,4	158
	600	79	1,2,3,4,5	2	3	1.1	97	1,2,3,5	1,2	1,2,3,4	200
Nakare	0	86	0	1	1	1.6	89	0	1	1	90
	100	49	0	1	1	1.3	88	0	1,2	1,2,4,5,6	75
	150	46	1,2,3,4	1	1	0.3	86	1,2,3	1,2	1,2,4,5,6	81
	200	8	1,2,3,4	1	1	0.5	80	1,2,3,4,5	1,2	1,2,4,5,6	71
	250	0	N/A	N/A	N/A	0.0	N/A	N/A	N/A	N/A	N/A
	300	0	N/A	N/A	N/A	0.0	N/A	N/A	N/A	N/A	N/A
Shindimba	0	88	0	1	1	1.9	95	0	1	1	70
	100	35	1,2,3,4	1	1	1.4	86	1,2,3	1,2	1,2,4,5,6	66
	150	37	1,2,3,4	1	1	0.8	93	1,2,3	1,2	1,2,4,5,6	65
	200	18	1,2,3,4	1	1	0.1	90	1,2,3,4,5	1,2	1,2,4,5,6	60
	300	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	400	0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

ES%, establishment in %; ABN, abnormalities observed, where 0 = normal, 1 = albino, 2 = leafy type, 3 = upright single stem, 4 = seedless pods, and 5 = short pods; FC, flower color, where 1 = white, 2 = purple. SC, seed color; where 1 = white, 2 = brown, 3 = red, 4 = cream, 5 = speckled, 6 = chocolate, 7 = light brown, 8 = black, 9 = mixed, 10 = dark brown. SYP, seed yield (gram/plant).

irradiation treatment and segregation. The following doses allowed successful selections of mutants during the M₃ to M₅: 300, 450, and 600 Gy (Bira), 100 and 150 Gy (Nakare), and 100 and 200 Gy (Shindimba). Surviving and phenotypically stable individuals were advanced at each selection generation at Omahenene and Bagani Research Stations. Qualitative traits had limited variation during M₃ to M₅ (Table 3). Bira mutants displayed purple FC irrespective of doses and test generations, while Nakare and Shindimba segregated for white and purple FC (Figures 3 and 5). Both Bira and Nakare mutants had straight PS similar to the controls. However, Shindimba segregants had straight and coiled pod types (Figure 5). Variable SCs including white, brown, red, cream, speckled, chocolate, light and dark brown, black and mixed were observed during the M₃ to M₅. Bira mutants had smooth SCT, while Nakare and Shindimba had mainly rough and smooth seed texture. Bushy, erect and spreading GHs were detected during the M₃ to M₅ (Figures 3 and 5).

Quantitative Traits Observed from M₃ to M₅

Quantitative traits of agronomic importance were measured during the M₃ to M₅ (Table 4). The percent seed emergence (ES%) reduced significantly with increased irradiation dose. Maximum seed germination was achieved 3 days after planting irrespective of irradiation doses (Tables 4–6). Shindimba mutants relatively flowered early (40 days) at the M₃ (Table 4).

At the M₄ a relatively shorter days to flowering (44 days) was recorded at 300 Gy (Table 5). Contrastingly, the number of days to flowering was 37 days at the M₅ at using 600 Gy (Table 6). Nakare derived mutants flowered relatively earlier

(10 days) at 100 Gy at the M₃ (Table 4). At the M₅ Nakare mutants recorded a minimum of 61 days to flowering at 0 and 150 Gy (Table 5). At the M₃, Shindimba mutants displayed a minimum of 15 and a maximum of 84 days to flowering at 200 and 100 Gy, respectively (Table 4). Nakare mutants recorded the lower days (25) for pod setting (DPS) at the M₃ when using 100 Gy. Comparatively, the higher number of DPS (98 days) was measured in Shindimba at 200 Gy. At the M₄ a minimum DPS of 48 days was recorded for Bira derivatives at 300 Gy. A maximum DPS of 86 days was recorded for Bira mutants at 400 Gy, Nakare at 100 and 150 Gy and Shindimba at 100 and 200 Gy (Table 5). At the M₅, Nakare mutants recorded the lower DPS (41 days) at 100 Gy, while Bira genotypes had the higher DPS of 88 days at 300 Gy, Table 6). During the M₃, Nakare mutants matured 32 days after planting at 100 Gy. At the same dose rate Shindimba displayed late maturity (98 days) at the M₃ (Table 4). During the M₄ Bira mutants matured earlier (54 days) at 450 Gy. Delayed maturity (115 days) were recorded for Nakare at 150 Gy and Shindimba at 100 and 200 Gy (Table 5). At the M₅ Bira measured early maturity (62 days) with the highest dose of 600 Gy. Interestingly, this genotype matured late (115 days) when subjected to irradiation dose of 300 Gy (Table 6). Nevertheless, Bira recorded lower NPP (1 pod/plant) at 600 Gy and higher (5 pods/plant) when irradiated at 300 Gy (Table 4). At the M₄, 1 pod/plant was recorded for Bira at 450 and 600 Gy and Shindimba at 200 Gy (Table 5).

At the M₃ the longer pod size measured at 23.5 cm was recorded for Nakare at 100 Gy (Table 4). At the M₄, Shindimba mutants resulted from 200 Gy measured longer pod size of



FIGURE 2 | Some common abnormalities at M₃ observed at Bagani Research Station. (A) spinach-like leaves, **(B)** Short-pods, **(C)** broad-dark leaves, **(D)** chlorophyll mutant, -single stem **(E,F)** observed at Omahenene research Station.

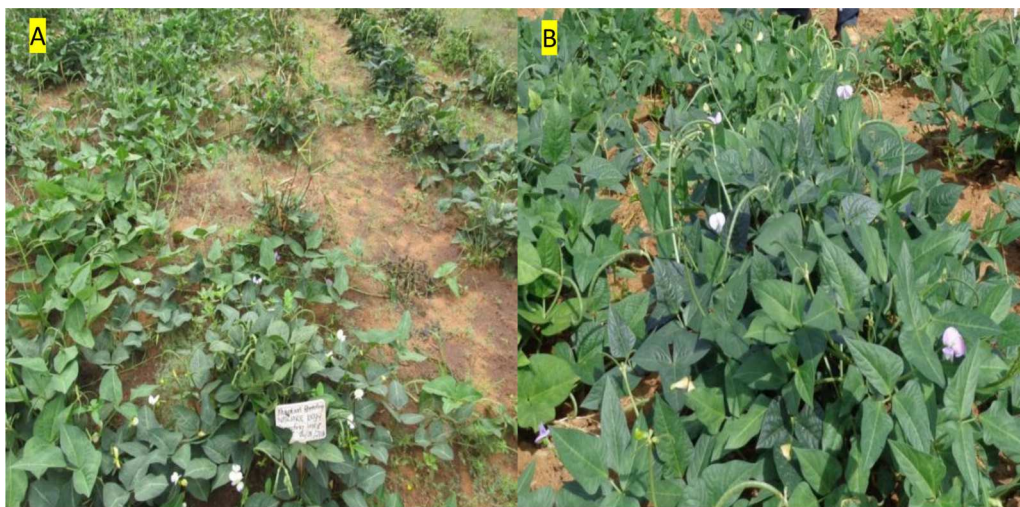


FIGURE 3 | Variation in flower color (A) white flower color, **(B)** purple flower and field plant stands of M₅ Nakare mutants observed at Omahenene Research Station in Namibia.

31 cm (Table 5). Bira mutants induced with 300 Gy produced longer pod size (30 cm) (Table 6). Relatively heavier pod size (4003 g/plant) was recorded for Bira at 300 Gy (Table 4). At the M₄, Bira had pod size measured at 325 g/plant at 300 Gy. Notably

this genotype had reduced pod weight (1 g/plant) at the highest irradiation dose (Table 5).

The NSP varied significantly between irradiation doses and genotypes. At the M₃, the highest number of seeds of 18.6/pod

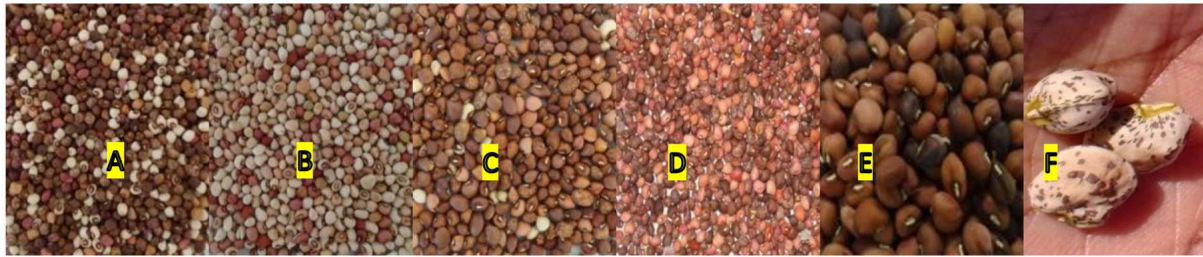


FIGURE 4 | Different M₃ seed colors (A–F) observed among all mutants at all locations.

TABLE 3 | Qualitative traits observed among the mutant lines at the M₃, M₄, and M₅ at Omahenene and Bagani Research Stations.

Genotype	Dose (Gy)	FC	PS	PC	SC	SCT	GH	PI
Bira	0	2	1	1	3	1	3	1
	300	2	1	1	1,2,3,4	1	1,2,3	1
	450	2	1	1	1,2,3,4	1	1,2,3	1
	600	2	1	1	1,2,3,4,5,6,7,8	1	1,2,3	1
Nakare	0	1	1	1	1	2	2	1
	100	1,2	1	1	1,2,3,6,7,9,10	1,2	1,2	1
	150	1,2	1	1	1,2,3,6,7,9,10	1,2	1,2	1
Shindimba	0	1	2	1	1	1,2	2	1
	100	1,2	1,2	1	1,2,3,7,9,10	1,2	1,2	1
	200	1,2	1,2	1	1,2,3,7,9,10	1,2	1,2	1

FC, flower color, where 1 = white, 2 = purple; PS, pod shape, where 1 = straight, 2 = coiled or curved; PC, pod color, where 1 = cream; SC = seed color, where 1 = white, 2 = brown, 3 = red, 4 = cream, 5 = speckled, 6 = chocolate, 7 = light brown, 8 = black, 9 = mixed, 10 = dark brown; SCT, seed coat texture, where 1 = smooth and 2 = rough; GH, growth habit, where 1 = bushy, 2 = erect and 3 = spreading; PI, pest infestation, where 0 = none, 1 = mild, and 2 = severe.



FIGURE 5 | Variation among Shindimba mutant lines. (A) coiled pods, (B) semi-coiled pods observed at Mannheim during the M₂ generation, (C) white flower with semi-coiled pods, and (D) Purple flowers observed at Omahenene during the M₅ generation.

TABLE 4 | Quantitative characteristics of M₃ cowpea mutant lines irradiated at different gamma radiation doses (Gy) in relation to their parental lines/control (Gy 0) observed at Bagani Research Station during 2011/2012 season.

Variety	Dose (Gy)	TNP	ES%		DG	DTF	DPS	DMT	NPP	PL	PW	NSP	HSW	SYP
Bira	0	330	100	Min	3	47	51	57	3.0	13.0	7.0	7.6	13.0	3.0
				Max	3	58	63	74	57.0	20.6	136	18.0	18.0	94.0
				Mean	3	50	55	64	28.6	15.1	64.9	14.5	15.1	40.9
	300	330	90.6	Min	3	40	49	52	4.0	10.4	3.0	9.0	9.0	1.0
				Max	3	80	86	92	5439.0	20.0	4003.0	19.0	25.0	3500.0
				Mean	3	52	58	63	322.6	16.9	211.9	10.8	14.4	136.6
	450	330	60.6	Min	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
				Max	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
				Mean	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	600	330	93	Min	3	46	50	57	1.0	5.0	1.0	1.8	4.0	1.0
				Max	3	59	63	74	120.0	21.3	218	18.6	21.0	142.0
				Mean	3	50	54	63	16.4.0	14.9	36.6	12.2	14.9	22.0
Nakare	0	330	45.5	Min	3	11	25	32	11.0	14.2	11.5	5.0	16.2	1.3
				Max	3	35	34	43	85.0	19.6	220.6	12.0	30.5	138.8
				Mean	3	20	29	36	51.2	15.8	106.1	7.7	25.3	59.2
	100	330	78.5	Min	3	10	25	32	3.0	6.8	2.8	2.0	11.4	0.7
				Max	3	37	62	65	125.0	23.5	327.8	17.2	32.1	191.2
				Mean	3	19	29	37	32.3	16.7	65.2	9.0	23.5	32.9
	150	330	60.3	Min	3	43	45	46	3.0	5.6	3.0	1.8	6.1	N/A
				Max	3	78	95	96	172.0	22.7	360.8	18.4	109	N/A
				Mean	3	52	57	66	39.0	14.2	67.6	10.2	15.8	N/A
Shindimba	0	330	69.7	Min	3	49	55	60	3.0	10.0	3.0	6.3	15.0	2.0
				Max	3	69	75	83	38.0	19.0	110.0	12.0	19.0	62.0
				Mean	3	57	62	70	23.0	16.0	44.7	9.5	17.5	22.2
	100	330	36	Min	3	49	54	60	3.0	10.0	3.0	5.8	14.0	1.0
				Max	3	84	90	98	40.0	21.0	110.0	16.6	25.0	118.0
				Mean	3	59	63	71	21.0	16.0	35.4	9.5	17.7	17.7
	200	330	69	Min	3	15	47	N/A	2.0	1.3	1.1	3.0	8.1	N/A
				Max	3	76	98	N/A	109.0	21.9	220.2	15.4	48.7	N/A
				Mean	3	62	68	N/A	32.0	14.6	55.5	7.9	20.8	N/A

TNP, total number of plants per plot; ES%, percentage establishment; DG, days to 50% germination; DTF, days to 50% flowering; DPS, days to 50% pod setting; DMT, days to 50% maturity; NPP, number of pods per plant; PL, pod length (in cm); PW, pod weight (gram), NSP, number of seeds per pod, HSW, 100 seed weight (gram), SYP, seed yield (gram/per plant), and N/A, data not available.

was recorded for Bira at 600 Gy and Nakare 150 Gy (**Table 4**). At the M₄ 19 seeds/pod was achieved in the mutants of Bira at 600 Gy and Shindimba at 100 Gy. At the M₅, mutants of Bira derived from 300 and 450 Gy and Nakare 150 Gy recorded 20 seeds/pod, the highest in this trial (**Table 5**). Hundred seed weight (HSW) at M₃ was relatively heavier measured at 109 g for Nakare mutants derived from 150 Gy (**Table 4**). At the M₄ the higher HSW (115 g) was recorded for Bira at 450 Gy (**Table 5**). During the M₅ Bira displayed higher HSW of 171 g at 450 Gy (**Table 6**). High seed yield per plant is an economic trait for cowpea growers. At M₃, higher seed yield of 3500 g per plant was recorded for Bira mutants derived from the mutagenic treatment of 300 Gy (**Table 4**).

During the M₄ generation Bira and Nakare mutants derived from 300 Gy and 100 Gy had a relatively higher seed yields of 287 and 199 g/plant, in that order (**Table 5**). At the M₅ generation Bira mutants yielded 570 g/plant, while Nakare had 298 g/plant when subjected to 450 Gy and 100 Gy, respectively (**Table 6**).

DISCUSSION

The present study revealed the important roles of induced mutations in cowpea breeding. It was evident from this study that increased Gy doses above 150 Gy can be lethal for the cowpea

TABLE 5 | Quantitative characteristics of M₄ cowpea mutant lines irradiated at different gamma radiation doses (Gy) in relation to their parental lines/control (Gy 0) observed at Omahene Research Station during 2012/2013 season.

Variety	Dose (Gy)	TNP	ES%		DG	DTF	DPS	DMT	NPP	PL	PW	NSP	HSW	SYP
Bira	0	330	69.7	Min	3	45	49	59	8.0	14.0	13.0	5.0	10.0	6.0
				Max	3	48	53	72	88.0	20.0	231.0	20.0	16.0	187.0
				Mean	3	46	50	68	31.0	17.7	86.2	14.2	13.0	53.4
	300	330	55.0	Min	3	44	48	66	2.0	9.0	4.0	6.0	5.0	1.0
				Max	3	51	55	74	97.0	21.0	325.0	18.0	79.0	287.0
				Mean	3	46	50	69	31.0	16.8	79.6	14.3	13.0	52.9
	450	330	85.7	Min	3	45	49	54	1.0	10.0	2.0	4.0	9.0	1.0
				Max	3	81	86	90	127.0	16.0	330.0	20.0	115.0	195.0
				Mean	3	49	54	60	26.0	15.8	50.0	17.0	15.0	30.0
	600	330	85.0	Min	3	46	50	57	1.0	6.0	1.0	2.0	4.0	1.0
				Max	3	59	63	74	124.	22.0	224.0	19.0	21.0	160.0
				Mean	3	50	55	63	18.4	16.3	41.6	13.0	15.1	25.1
Nakare	0	330	42.0	Min	3	61	66	96	5.0	11.0	9.0	4.0	17.0	6.0
				Max	3	74	78	110	32.0	19.0	85.0	17.0	26.0	62.0
				Mean	3	70	75	104	14.0	15.1	29.8	9.8	22.8	22.8
	100	330	56.0	Min	3	61	66	86	2.0	10.0	4.0	5.0	5.0	2.0
				Max	3	78	86	113	70.0	21.0	227.0	14.0	59.0	199.0
				Mean	3	71	76	103	15.0	15.5	35.3	10.0	21.4	26.7
	150	330	88.8	Min	3	61	67	86	2.0	8.0	3.0	3.0	3.0	1.0
				Max	3	79	86	115	85.0	26.0	287.0	18.0	40.0	131.0
				Mean	3	71	76	103	21.9	16.5	48.0	11.0	17.3	33.5
Shindimba	0	330	93.9	Min	3	42	68	72	7.0	7.0	7.0	3.0	12.0	3.0
				Max	3	75	78	85	44.0	29.0	123.0	13.0	30.0	91.0
				Mean	3	71	72	76	20.0	13.0	33.0	7.4	20.1	25.1
	100	330	82.4	Min	3	42	66	87	2.0	7.0	3.0	3.0	10.0	2.0
				Max	3	80	86	115	63.0	23.0	130.0	19.0	30.0	91.0
				Mean	3	71	76	104	16.7	13.8	27.7	8.1	19.8	20.9
	200	330	68.5	Min	3	62	66	94	1.0	9.0	3.0	3.0	6.0	2.0
				Max	3	80	86	115	59.0	31.0	123.0	16.0	30.0	91.0
				Mean	3	72	76	104	15.5	13.0	26.5	8.1	18.4	19.3

TNP, total number of plants per plot; ES%, percentage establishment; DG, days to 50% germination; DTF, days to 50% flowering; DPS, days to 50% pod setting; DMT, days to 50% maturity; NPP, number of pods per plant; PL, pod length (in cm); PW, pod weight (gram), NSP, number of seeds per pod, HSW, 100 seed weight (gram), SYP, seed yield (gram/per plant).

breeding line such as Nakare, while a dose above 200 Gy is lethal for the breeding line Shindimba. Other authors have reported the negative effects of increased mutagenic doses affecting various crops' establishment and survival for breeding (Mba et al., 2009).

The present study showed the presence of clear phenotypic differences among the tested mutant lines when compared to their respective controls. Overall, increased irradiation dose was associated with visual phenotypic mutants. Mutants displayed visual phenotypic differences including chlorophyll, leaf, upright single stem, pod, and seed during the M₂ to M₅ generations. Chlorophyll mutants observed were plants with yellow and striped leaves, albinos or yellow to pale leaf and stem pigmentations. Virescence mutants showed broad pale green leaf

with its margin resembling a spinach leaf (**Figure 2**). According to Girija and Dhanavel (2009) and Maluszynski et al. (2009), the appearance of chlorophyll defects is a good indicator of genetic action of the mutagen. Singh et al. (2013) reported that increased Gy doses provided higher frequency of chlorophyll mutants in cowpea when compared to other mutagens such as EMS. Girija and Dhanavel (2009) outlined the effectiveness and efficiency of mutagens for selection of mutants with economic traits. The authors suggested that for effective phenotypic selection the mutation treatment should not yield unintended damages including chromosomal aberrations, physiological and toxic effects, which reduce cell survival and ultimately eliminate the mutation. Despite its negative effects on the early stages of

TABLE 6 | Quantitative characteristics of M₅ cowpea mutant lines irradiated at different gamma radiation doses (Gy) in relation to their parental lines/control observed at Omahene Research Station during 2013/2014 season.

Variety	Dose (Gy)	TNP	ES%		DG	DTF	DPS	DMT	NPP	PL	PW	NSP	HSW	SYP
Bira	0	330	97.0	Min	3	68.0	73.0	98.0	7.0	16.0	15.0	6.0	9.0	10.0
				Max	3	83.0	88.0	115.0	66.0	27.0	155.0	18.0	16.0	115.0
				Mean	3	74.0	78.0	102.0	40.7	21.0	88.9	14.3	12.9	61.9
	300	330	77.6	Min	3	64.0	69.0	98.0	3.0	13.0	6.0	5.0	4.0	3.0
				Max	3	83.0	88.0	115.0	150.0	30.0	325.0	20.0	29.3	213.0
				Mean	3	73.0	78.0	102.0	30.9	21.0	66.6	14.2	12.7	47.0
	450	330	85.8	Min	3	42.0	46.0	66.0	2.0	13.0	6.0	5.0	11.0	3.0
				Max	3	58.0	69.0	76.0	233.0	20.0	659.0	20.0	171.0	570.0
				Mean	3	47.0	51.0	70.0	31.6	17.7	81.5	14.9	16.2	60.0
	600	330	78.5	Min	3	37.0	42.0	62.0	1.0	9.0	1.0	3.0	4.0	2.0
				Max	3	56.0	61.0	81.0	78.0	27.0	276.0	18.0	19.0	157.0
				Mean	3	46.0	50.3	68.6	19.9	16.3	43.4	13.2	12.4	28.1
Nakare	0	330	42.0	Min	3	41.0	45.0	65.0	47.0	16.0	76.0	13.0	1.0	51.0
				Max	3	53.0	57.0	79.0	46.0	21.0	72.0	12.0	298.0	51.0
				Mean	3	47.0	50.8	69.0	45.0	16.8	70.8	12.3	62.8	50.5
	100	330	56.4	Min	3	37.0	41.0	58.0	1.0	9.0	1.0	1.0	1.0	1.0
				Max	3	57.0	60.0	80.0	144.0	23.0	375.0	18.0	81.0	298.0
				Mean	3	46.0	49.5	66.0	39.0	17.7	86.8	12.0	18.1	62.8
	150	330	59.7	Min	3	42.0	46.0	64.0	1.0	10.0	3.0	5.0	6.0	2.0
				Max	3	53.0	57.0	73.0	110.0	28.0	317.0	20.0	82.0	209.0
				Mean	3	46.0	50.0	68.0	29.0	21.0	70.6	12.5	18.4	45.6
Shindimba	0	330	93.9	Min	3	50.0	54.0	66.0	2.0	8.0	5.0	3.0	2.0	2.0
				Max	3	55.0	59.0	73.0	88.0	20.0	127.0	11.0	29.0	90.0
				Mean	3	52.3	56.0	69.0	33.8	13.6	64.0	8.0	22.0	42.0
	100	330	86.7	Min	3	42.0	46.0	60.0	1.0	8.0	1.0	1.0	2.0	1.0
				Max	3	67.0	70.0	89.0	122.0	25.0	392.0	18.0	29.0	208.0
				Mean	3	50.5	54.2	70.0	35.5	14.0	75.0	8.0	21.5	50.0
	200	330	83.6	Min	3	44.0	48.0	65.0	1.0	7.0	1.0	1.0	6.0	1.0
				Max	3	67.0	71.0	91.0	89.0	29.0	193.0	18.0	25.0	93.0
				Mean	3	52.0	56.0	73.0	20.0	16.0	39.0	9.0	16.0	25.0

TNP, total number of plants per plot; ES%, percentage establishment; DG, days to 50% germination; DTF, days to 50% flowering; DPS, days to 50% pod setting; DMT, days to 50% maturity; NPP, number of pods per plant; PL, pod length (in cm); PW, pod weight (gram), NSP, number of seeds per pod, HSW, 100 seed weight (gram), SYP, seed yield (gram/per plant).

crop growth, chlorophyll mutants are important in mutation breeding programs. Tulmann Neto et al. (2011) reported that the chlorophyll mutants were used in evaluation of the genetic effects and sensitivity of various mutagens on crops. These results are in agreement with Goyal and Khan (2010) whose studies indicated that the incidence of chlorophyll mutants were higher with increased Gy doses in earlier selection generations.

In the present study, mutants at the M₂ were genetically diverse owing to phenotypic segregation. The genetic diversity assessed in these mutants were tall/dwarf plant heights, early/late maturity, leaf shapes, branching habit, GH, PS, FC, SC and texture, seed weight and yield (Tables 4–6). Both the qualitative and quantitative parameters measured in the study were useful

for selection of cowpea mutants. According to Maluszynski et al. (2009), induced genetic polymorphism among initial cells of the sporogenic layer influences the segregation ratio in the M₂ generation. However, mutations of cells of somatic tissues are not transferred to the next generation. Gnanamurthy et al. (2012) stipulated that easily detectable mutants characteristics are phenotypically visible and morphologically distinct with qualitatively inherited genetic changes. These changes occur due to the effect of few major genes or oligogenes yielding macro mutations. In this study, some macro mutations observed were the changes in flower and SC. Micro mutations are the result of polygenes each with minor genetic effect showing quantitative inheritance. The effect and inheritance of minor genes is detected

using quantitative genetic parameters and statistical methods (Singh et al., 2006). In the current study, short plant height and one seed per pod mutants were recorded in all the breeding lines mostly at the M₃ generation. Single seeded pods were also reported by Girija and Dhanavel (2009).

In the present study, other main phenotypic changes observed were increased NMB especially in mutants with spreading GH. Mutants with bushy GH had reduced number of branches per plant. These characters are indicated to be associated with some physiological properties of the plant including leaf senescence and indeterminate GH (Hall, 2004; Martins et al., 2014). It is reported that characteristics altered through mutation breeding can be combined through the conventional breeding to improve crop performance and drought adaptation (Ehlers and Hall, 1997). The present study found that Nakare mutants had a maximum of 23 main branches per plant, while the comparative control had nine main branches (Table 3). According to previous studies (Singh et al., 2003, 2013), the spreading and semi-spreading cowpea types yielded less grain and more fodder when planted in closer spaced rows. The present study found that mutation treatment did not significantly affect the number of days taken to germination, hence all the breeding lines germinated 3 days after planting (Tables 3–6). The mutation treatment had positive effect on the number of days taken to 50% flowering whereby some of the breeding lines flowered 11 days before the control. Bira mutants subjected to irradiation of 300 Gy flowered 80 days after planting (Table 3). Maluszynski et al. (2009) suggested that a high dose of a mutagen should yield delayed maturity. Dhanavel et al. (2008) reported that mutagenesis resulted into variation in plant development including the number of days taken to maturity. According to

Singh et al. (2003), these variations are important to the farmers and the breeders allowing choices of planting time. The breeder will have a choice from a larger breeding stock for various breeding traits and purposes.

Significant observations made in the present study were increased PL and seed yield measured during the M₃ to M₅ in all the breeding lines. Goyal and Khan (2010) reported that mutations caused increased PL in some of the cowpea lines. Pod size may contribute to increased seed yield. The number of grains per pod increases with increased PL though this may be associated with reduced total biomass (Singh et al., 2003). Other major effects of the mutation observed in the present study were the range of variations in SC. A mosaic of SCs were noted including white, brown, chocolate, red, speckled, cream, and black. Dhanavel et al. (2008) reported various SCs due to mutational events. The present findings suggested that the NMB per plant, NPP, number of grains per pod, 100-seed weight and seed yield per plant reduced significantly with increased concentration of irradiation doses. These findings are in agreement to the studies of Girija and Dhanavel (2009), who reported that mutagenesis is associated with negative and positive phenotypic effects for selection.

The present study demonstrated that most characters of cowpea which are of interest to plant breeders can be altered through mutations using the gamma irradiation technique. Furthermore, new plant attributes were created in the high yielding and well adapted local cowpea varieties. Various pests were observed on mutant cowpea during this study Figure 6. Therefore, there is a need to breed for insect pest tolerance in cowpea. Timko et al. (2007) suggested that the future of cowpea improvement programs should focus on breeding



FIGURE 6 | Common insect pests (A) Spiny brown bugs *Clavigralla* sp., (B) Coreid bug *Anoplocnemis curvipes*, (C) Aphids *Aphis craccivora* Koch and Blister (D) Beetle *Mylabris phalerata* observed among the M₅ mutants at Bagani, and Omahenene Research Stations concurrently.

for pests and diseases resistance and other desirable traits such as early maturity, photoperiod insensitivity, suitable plant type, seed quality and yield. Overall, the present study made extensive phenotypic selections of mutants from the M₂ to M₅ generations and identified promising genotypes. The selected mutants' are recommended for adaptability and stability tests across representative agro-ecologies for large-scale production or breeding in Namibia or similar environments. The novel cowpea genotypes selected through the study are valuable genetic resources for genetic enhancement and breeding.

AUTHOR CONTRIBUTIONS

LH and HS designed the research. The experiments were carried out by LH under the supervision of HS. The manuscript was prepared by LH, HS, and HG.

REFERENCES

- Adekola, O., and Oluleye, F. (2007). Induction of genetic variation in Cowpea (*Vigna unguiculata* L. Walp.) by gamma irradiation. *Asian J. Plant Sci.* 6, 869–873. doi: 10.3923/ajps.2007.869.873
- Bressani, R. (1985). "Nutritive value of cowpea," in *Cowpea: Research, Production and Utilization*, eds S. R. Singh and K. O. Rachie (New York, NY: John Wiley and Sons).
- Ceccarelli, S., and Grando, S. (2007). Decentralized-participatory plant breeding: an example of demand driven research. *Euphytica* 155, 349–360. doi: 10.1007/s10681-006-9336-8
- De Ronde, J. A., and Spreeth, M. H. (2007). Development and evaluation of drought resistant mutant germ-plasm of *Vigna unguiculata*. *Water SA* 33, 381–386.
- Dhanavel, D., Pavada, P., Mullainathan, L., Mohana, D., Raju, G., and Girija, M. (2008). Effectiveness and efficiency of chemical mutagens in cowpea (*Vigna unguiculata* (L.) Walp.). *Afr. J. Biotechnol.* 7, 4116–4117.
- Ehlers, J. D., and Hall, A. E. (1997). Cowpea (*Vigna unguiculata* L. Walp.). *Field Crops Res.* 53, 187–204. doi: 10.1016/S0378-4290(97)00031-2
- Ghaly, A. E., and Alkoai, F. N. (2010). Extraction of protein from common plant leaves for use as human food. *Am. J. Appl. Sci.* 7, 331–342. doi: 10.3844/ajassp.2010.331.342
- Girija, M., and Dhanavel, D. (2009). Mutagenic effectiveness and efficiency of gamma rays, ethyl methane sulphonate and their combined treatments in cowpea (*Vigna unguiculata* L. Walp.). *Global J. Mol. Sci.* 4, 68–75.
- Gnanamurthy, S., Mariyammal, S., Dhanavel, D., and Bharathi, T. (2012). Effect of gamma rays on yield and yield components characters R3 generation in cowpea (*Vigna unguiculata* (L.) Walp.). *Int. J. Res. Plant Sci.* 2, 39–42.
- Goyal, S., and Khan, S. (2010). Induced mutagenesis in Urd bean (*Vigna mungo* (L.) Hepper): a review. *Int. J. Bot.* 6, 194–206. doi: 10.3923/ijb.2010.194.206
- Hall, A. E. (2004). Breeding for adaptation to drought and heat in cowpea. *Eur. J. Agron.* 21, 447–454. doi: 10.1016/j.eja.2004.07.005
- Horn, L., and Shimelis, H. (2013). Radio-sensitivity of selected cowpea (*Vigna unguiculata*) genotypes to varying gamma irradiation doses. *Sci. Res. Essays* 8, 1991–1997.
- International Institute for Tropical Agriculture [IITA] (2004). *IITA, Annual Report 2004*. Available at: http://www.iita.org/c/document_library/get_file?uuid=9b4353cd-189f-4a1c-a224-85ee65db18d8&groupId=25357 (accessed October 23, 2015).
- Maluszynski, M. (2001). Officially released mutant varieties – the FAO/IAEA database. *Plant Cell Tissue Organ Cult.* 65, 175–177. doi: 10.1023/A:1010652523463
- Maluszynski, M., Szarejko, I., Bhatia, C. R., Nichterlein, K., and Lagoda, P. J. L. (2009). "Methodologies for generating variability. Part 4: Mutation techniques," in *Plant Breeding and Farmer Participation*, eds S. Ceccarelli and E. Weltzien (Rome: Food and Agriculture Organization of the United Nations), 159–194.
- Martins, C. M., Lawlor, D. W., Quilambo, O. A., and Kunert, K. J. (2014). Evaluation of four Mozambican cowpea landraces for drought tolerance. *S. Afr. J. Plant Soil* 31, 87–91. doi: 10.1080/02571862.2014.907453
- Mba, C., Afza, R., Bado, S., and Jain, S. M. (2010). "Induced mutagenesis in plants using physical and chemical agents," in *Plant Cell Culture: Essential Methods*, eds M. R. Davey and P. Anthony (Chichester: John Wiley and Sons), 111–130.
- Mba, C., Afz, R., Jankowicz-Cieslak, J., Bado, S., Matijevic, M., Huynh, O., et al. (2009). "Enhancing genetic diversity through induced mutagenesis in vegetatively propagated plants," in *Proceedings International Symposium Induced Mutations in Plants: Induced Plant Mutations in the Genomics Era, 11–15 August 2008*, Vienna.
- Ng, O., and Marachel, R. (1985). "Cowpea taxonomy, origin, and germplasm," in *Cowpea Research, Production and Utilization*, eds S. R. Singh and K. O. Rachie (Chichester: John Wiley and Sons), 13–21.
- Nielen, S. (2004). FAO/IAEA mutant variety database. Mutation plant breeding, IAEA/FAO Vienna. *Plant Cell Tissue Organ Cult.* 65, 175–177.
- SAS (2002). *SAS Institute Inc., SAS 9.1.3 Help and Documentation*. Cary, NC: SAS Institute Inc.
- Singh, B. B., Ajeigbe, H. A., Tarawali, S. A., Fernandez-Rivera, S., and Abubakar, M. (2003). Improving the production and utilization of cowpea as food and fodder. *Field Crops Res.* 84, 169–177. doi: 10.1016/S0378-4290(03)00148-5
- Singh, D. P., Sharma, S. P., Lal, M., Ranwah, B. R., and Sharma, V. (2013). Induction of Genetic variability for polytraits through physical and chemical mutagens in cowpea (*Vigna unguiculata*). *Legume Res.* 36, 10–14.
- Singh, V., Ramkrishna, K., and Arya, R. K. (2006). Induced chemical mutagenesis in cowpea [*Vigna unguiculata* (L.) Walp.]. *Indian J. Genet. Plant Breed.* 66, 312–315.
- Steele, W. M. (1972). *Cowpea in Africa*. Doctoral thesis, University of Reading, Reading.
- Timko, M. P., Ehlers, J. D., and Roberts, P. A. (2007). "Cowpea, pulses, sugar and tuber crops," in *Genome Mapping and Molecular Breeding in Plants*, ed C. Kole (Heidelberg: Springer-Verlag Berlin), 49–67.
- Tshilenge-Lukanda, L., Funny-Biola, C., Tshiyoyi-Mpunga, A., Mudibu, J., Ngoie-Lubwika, M., and Mukendi-Tshibingu, R. (2012). Radio-sensitivity of some groundnut (*Arachis hypogaea* L.) genotypes to gamma irradiation: indices for use as improvement. *Br. J. Biotechnol.* 3, 169–178. doi: 10.9734/BBJ/2012/1459
- Tulmann Neto, A., Ando, A., Figueira, A., Latado, R. R., dos Santos, P. C., Correa, L. S., et al. (2011). Genetic improvement of crops by mutation techniques in Brazil. *Plant Mutat. Rep.* 2, 24–37.
- Udensi, O., Arong, G., Obu, J., Ikpeme, E., and Ojobe, T. (2012). Radio-sensitivity of some selected landraces of pulses to gamma irradiation: indices for use as improvement and preservation techniques. *Am. J. Exp. Agric.* 2, 320–335.

FUNDING

This work was supported by funds from the International Atomic Energy Agency (IAEA) and Ministry of Agriculture, Water and Forestry of Namibia.

ACKNOWLEDGMENTS

The authors wish to thank the University of KwaZulu-Natal and the Ministry of Agriculture, Water and Forestry (MAWF) of the Government of Namibia for the overall research support to the first author. The support from the technical team (Loide Aron, Rose-Marrie Hukununa, Kangumba Annethe, and Nghishekwa Alfeus) is highly appreciated.

Wani, M. R. (2006). Estimates of genetic variability in mutated populations and the scope of selection for yield attributes in *Vigna radiata* (L.) Wilczek. *Egypt. J. Biol.* 8, 1–6.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Horn, Ghebrehewot and Shimelis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



EcoTILLING-Based Association Mapping Efficiently Delineates Functionally Relevant Natural Allelic Variants of Candidate Genes Governing Agronomic Traits in Chickpea

OPEN ACCESS

Edited by:

Nicolas Rispaill,
Institute for Sustainable Agriculture -
Consejo Superior de Investigaciones
Científicas, Spain

Reviewed by:

Milind Ratnaparkhe,
Directorate of Soybean Research,
India
Samira Mafi Moghaddam,
North Dakota State University, USA

*Correspondence:

Swarup K. Parida
swarup@nipgr.ac.in;
swarupdbt@gmail.com

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 28 November 2015

Accepted: 22 March 2016

Published: 19 April 2016

Citation:

Bajaj D, Srivastava R, Nath M,
Tripathi S, Bharadwaj C,
Upadhyaya HD, Tyagi AK and
Parida SK (2016) EcoTILLING-Based
Association Mapping Efficiently
Delineates Functionally Relevant
Natural Allelic Variants of Candidate
Genes Governing Agronomic Traits in
Chickpea. *Front. Plant Sci.* 7:450.
doi: 10.3389/fpls.2016.00450

Deepak Bajaj¹, Rishi Srivastava¹, Manoj Nath², Shailesh Tripathi³, Chellapilla Bharadwaj³, Hari D. Upadhyaya⁴, Akhilesh K. Tyagi¹ and Swarup K. Parida^{1*}

¹ Govt. of India, Plant Genomics and Molecular Breeding Lab, Department of Biotechnology, National Institute of Plant Genome Research, New Delhi, India, ² National Research Centre on Plant Biotechnology, New Delhi, India, ³ Division of Genetics, Indian Agricultural Research Institute, New Delhi, India, ⁴ International Crops Research Institute for the Semi-Arid Tropics, Patancheru, India

The large-scale mining and high-throughput genotyping of novel gene-based allelic variants in natural mapping population are essential for association mapping to identify functionally relevant molecular tags governing useful agronomic traits in chickpea. The present study employs an alternative time-saving, non-laborious and economical pool-based EcoTILLING approach coupled with agarose gel detection assay to discover 1133 novel SNP allelic variants from diverse coding and regulatory sequence components of 1133 transcription factor (TF) genes by genotyping in 192 diverse *desi* and *kabuli* chickpea accessions constituting a seed weight association panel. Integrating these SNP genotyping data with seed weight field phenotypic information of 192 structured association panel identified eight SNP alleles in the eight TF genes regulating seed weight of chickpea. The associated individual and combination of all SNPs explained 10–15 and 31% phenotypic variation for seed weight, respectively. The EcoTILLING-based large-scale allele mining and genotyping strategy implemented for association mapping is found much effective for a diploid genome crop species like chickpea with narrow genetic base and low genetic polymorphism. This optimized approach thus can be deployed for various genomics-assisted breeding applications with optimal expense of resources in domesticated chickpea. The seed weight-associated natural allelic variants and candidate TF genes delineated have potential to accelerate marker-assisted genetic improvement of chickpea.

Keywords: allele, association mapping, chickpea, EcoTILLING, seed weight, SNP, transcription factor

INTRODUCTION

Allele mining is an efficient strategy to unlock a wealth of largely untapped natural and functional allelic variation/diversity existing within wild and cultivated genetic resources for crop genetic enhancement, thereby improving the productivity and sustainability of global agriculture. The vast available germplasm (core and mini-core) repositories and different recently developed high-throughput array-based next-generation sequencing (NGS) and sequence-based marker genotyping strategies are found expedient in large-scale mining and genotyping of genome/gene-based SNP (single nucleotide polymorphism) alleles among these germplasm accessions for driving genomics-assisted crop improvement through genetic and association mapping. The allele mining strategies commonly adopted in laboratories equipped with advanced infrastructural facilities (like high-throughput genotyping platforms and modern computational genomics tools), require prior information of SNP alleles (nature/types and flanking sequences) for their discovery, validation and genotyping in the targeted gene/genomic regions of multiple crop accessions. Conversely, EcoTILLING (Ecotype Targeting Induced Local Lesions IN Genomes), a rapid, inexpensive and well-established allele mining approach is found much proficient in large-scale mining and high-throughput genotyping of novel natural and functional allelic variants (without prior knowledge of SNP alleles) of known and candidate genes related to useful agronomic traits in diverse crop germplasm accessions (McCallum et al., 2000; Comai et al., 2004; Till et al., 2006, 2007, 2010; Raghavan et al., 2007; Wang et al., 2010; Xia et al., 2012). The implication of EcoTILLING to identify potential novel functional alleles in the known and candidate genes/transcription factors (TFs) regulating qualitative and quantitative agronomic traits by association/genetic mapping is well-documented for expediting the genetic enhancement of crop plants (Mejlhede et al., 2006; Barkley and Wang, 2008; Ibiza et al., 2010; Negrao et al., 2011; Yu et al., 2012; Frerichmann et al., 2013).

EcoTILLING usually employs a mismatch-specific CEL-I nuclease to cleave the PCR amplified fragments at the site of heteroduplex formation involving nucleotide (SNP-allelic) polymorphism. Most of the EcoTILLING studies utilize the advanced genotyping platforms (LICOR NEN Model 4300 DNA Analyzer, Transgenomic WAVE-HS denaturing high performance liquid chromatography, ABI 377 sequencer and eGene capillary electrophoresis systems) for efficient resolution of fluorescent dye (IRDye 700/800 and SYBR green)-labeled CEL-I cleaved heteroduplex PCR amplified fragments. Consequently, these efforts led to the discovery and genotyping of novel potential alleles specifically derived from the trait-associated known and candidate genes in natural population of diverse crop plants (Perry et al., 2003; Caldwell et al., 2004; Comai et al., 2004; Henikoff et al., 2004; Yang et al., 2004; Suzuki et al., 2005). The added-advantage of agarose gel-based EcoTILLING vis-à-vis the commonly utilized LICOR genotyper for large-scale mining and genotyping of allelic variants in accessions exhibiting low level polymorphism, is well-demonstrated in many crop plants (Raghavan et al., 2007;

Negrao et al., 2011; Yu et al., 2012). This is merely because efficacy of an agarose gel-based EcoTILLING approach in precise resolution of unlabeled CEL I-cleaved heteroduplex PCR amplified fragments by a simpler, economical and time-saving agarose gel-based detection assay as compared to a standard EcoTILLING method that requires labeled CEL I-cleaved heteroduplex PCR amplicons for resolution in a LICOR genotyper. The broader utility and deployment of this agarose gel-based EcoTILLING approach in manifold large-scale genotyping applications is well-documented by the research laboratories with minimal resources (Raghavan et al., 2007; Negrao et al., 2011; Yu et al., 2012). This includes understanding the natural allelic diversity, population genetic structure and domestication pattern among accessions, molecular mapping and genetic association analysis for identification of potential molecular tags like alleles and genes/QTLs (quantitative trait loci) governing vital agronomic traits and marker-assisted breeding for selecting desirable accessions for crop genetic improvement.

Chickpea, a member of genus *Cicer*, is rich in cultivated and wild germplasm resource (core/mini-core collections) with a wealth of trait diversity (Upadhyaya and Ortiz, 2001; Upadhyaya et al., 2001, 2002). More in-depth characterization of these core/mini-core germplasm resources at both genotypic and phenotypic level for diverse important abiotic/biotic stress tolerance and yield/quality component traits is essential to discover and deploy valuable alleles and allelic combinations scanned from these germplasm accessions, more effectively for genetic improvement of chickpea (Upadhyaya et al., 2008, 2011; Varshney et al., 2013a; Saxena et al., 2014a,b; Bajaj et al., 2015). The existing diverse germplasm collections are thus “gold mines” for analysis of functional as well as natural allelic variation/diversity in the known and candidate genes controlling important agronomic traits of chickpea. Considering the importance of allele mining in crop genetic enhancement, EcoTILLING can be employed in multiple cultivated (*desi* and *kabuli*) and wild chickpea accessions for identifying novel functional/natural allelic variants in the candidate and known genes associated with multiple traits of agricultural importance in chickpea.

The agarose gel-based EcoTILLING strategy mostly utilizes pooling of genomic DNA isolated from two diverse accessions rather than multiple accessions for robust mining and genotyping of alleles in the view of anticipating more allelic variations between distant accessions of crop plants (Raghavan et al., 2007). However, the level of allelic variation and diversity captured specifically from different sequence components of genes/genomes among germplasm accessions of chickpea is known to be very low due to its narrow genetic base and extensive domestication bottlenecks as compared to other crop plants (Abbo et al., 2003, 2005; Berger et al., 2003, 2005; Singh et al., 2008; Toker, 2009; Jain et al., 2013; Varshney et al., 2013b; Saxena et al., 2014a). Therefore, in a diploid self-pollinated crop species like chickpea with a lower occurrence of SNP-allelic variations, the agarose gel-based EcoTILLING approach can easily be expanded to multiple accessions regardless of selecting only two accessions for DNA pooling in allele mining. Consequently, the efficient resolution and estimation of allelic

variants scanned from the pooled DNA of multiple chickpea accessions will be relatively convenient, even in a low-resolution agarose gel than that of other crop species exhibiting higher allelic polymorphism. Such a strategy of multiple accessions pooling-based EcoTILLING coupled with agarose gel detection approach has been found beneficial for various high-throughput allele mining and large-scale genotyping applications, including genetic and association mapping of alleles/genes (TFs) regulating drought and salinity stress tolerance traits in rice (Negrao et al., 2011; Yu et al., 2012). Henceforth, the utilization of this multiple accessions-pooling agarose gel-based EcoTILLING approach can certainly accelerate the process of rapid selection of informative SNP alleles/markers as well as identification of accessions exhibiting higher allelic variations for their robust genotyping at a genome-wide scale. This strategy will be thus useful for various high-throughput genetic analysis in chickpea with sub-optimal use of resources. A large-scale novel as well as functional allelic genotyping information cataloged from diverse germplasm (core/mini-core) accessions and bi-parental mapping populations by use of agarose gel-based EcoTILLING assay can serve as a vital resource for trait association and genetic mapping. This will be helpful to identify favorable natural allelic variants undergoing selection during the course of domestication in *desi*, *kabuli* and wild accessions that are adapted to diverse agro-climatic conditions for genomics-assisted crop improvement of chickpea.

In light of the above, the present study employed a simpler non-laborious and rapid yet cost-effective agarose gel-based EcoTILLING assay (Figure 1) for high-throughput mining of natural allelic variants derived from diverse coding and non-coding regulatory sequence components of 1133 TF genes by genotyping in 192 core/mini-core germplasm accessions constituting a seed weight association panel. As a proof of concept, the high-throughput genotyping data of 1133 TF gene-derived SNPs was correlated with seed weight field phenotypic information of the 192 accessions to delineate functionally relevant natural allelic variants in the candidate TF genes regulating 100-seed weight in chickpea.

AGAROSE GEL-BASED ECOTILLING AIDS IN MINING OF NOVEL NATURAL AND FUNCTIONAL ALLELIC VARIANTS IN CHICKPEA

For large-scale mining and genotyping of gene-based SNP alleles by EcoTILLING in chickpea, a set of 1248 TF genes annotated from *desi* and *kabuli* genomes were acquired. The selected TFs include 819 *desi* and *kabuli* TF genes and 429 TF-encoding transcripts of *desi* accession (ICC 4958), which were specifically selected from the previous studies of Jhanwar et al. (2012) and Kujur et al. (2015), respectively based on the presence of at least one SNP in the CDS (coding sequence) and regulatory sequences of these genes. The multiple forward and reverse primer combination (at least two primer-pairs per TF gene) with expected amplification product size of 1000–1500 bp (per primer) targeting the diverse CDS and 2000-bp

upstream/downstream regulatory regions (URRs/DRRs) of 1248 TF genes were designed. The amplification of each target gene regions was optimized (specifically the annealing temperature) with different combination of primer-pairs using the genomic DNA of one *desi* chickpea accession (ICC 4958) as per the detail PCR protocol described by Jhanwar et al. (2012) and Kujur et al. (2013). Based on these analyses, 1890 (75.7% of 2496 primer-pairs designed in total) primers designed from the 1133 TF genes exhibited reproducible single amplicons (by eliminating the non-specific amplified fragments and duplicate loci) in ICC 4958 using 2.5% agarose gel (Figure 2A). The fragments amplified specific to diverse coding (CDS) and non-coding URR/DRR sequence components of TF genes using the optimized primers were further assayed through agarose gel-based EcoTILLING approach for allele mining.

To access the potential of EcoTILLING in large-scale mining and high-throughput genotyping of TF gene-derived SNP alleles, 192 *desi* and *kabuli* chickpea core/minicore germplasm accessions were selected (Table S1) from a 100-seed weight (SW) specific association panel (244 accessions) as constituted previously by Kujur et al. (2014). The high-quality genomic DNA isolated from these 192 accessions was quantified to equal concentration of 1 ng/μl. The bi-dimensional pooling of the uniformly quantified genomic DNA of 192 accessions was performed in two of each 96-well PCR plate to constitute eight micropools and one superpool (per plate) according to Tsai et al. (2011) (Figure 1). The genomic DNA of each of these pools was contrasted with that of ICC 4958 individually with a 1:1 ratio and further PCR amplified with the 1890 optimized primer-pairs designed from the 1133 TF genes (as per aforementioned methods). The amplified PCR product from each pool was denatured and renatured for homoduplex/heteroduplex formation and digested with CEL I-based SNIperase-L enzyme (FRONTIER GENOMICS, Alaska, USA) following the detail instructions of manufacturer (FRONTIER GENOMICS; Figure 1). The purified CEL I cleaved homo/heteroduplex PCR products of each TF gene amplified from the pools were resolved in 2.5% agarose gel as per the EcoTILLING approach documented by Raghavan et al. (2007) (Figure 1). The individual accession exhibiting putative mutations (SNP allelic variants) was screened from the pools by accessing the digestion pattern of all 1133 TF genes in the row and/or column-wise de-multiplexed genomic DNA following the aforesaid agarose gel-based EcoTILLING method (Figures 2B,C). To ascertain the putative mutations (SNP allelic variants) discovered in the TF genes among accessions constituting the pools, the PCR products of corresponding genes amplified from the pools/accessions were sequenced by an automated 96 capillary ABI 3730xl DNA Analyzer (Applied Biosystems, USA) (Figure 2D). The SNP allelic variants were detected by aligning and comparing the multiple high-quality gene sequences among accessions following Kujur et al. (2013). The above-said analysis of allele mining and genotyping by agarose gel-based EcoTILLING led to discover 1133 SNP allelic variants from the diverse coding and non-coding regulatory sequence components of 1133 TF genes (Table S2). Of these, 406 (35.8%) and 702 (62.0%) SNP alleles exhibited

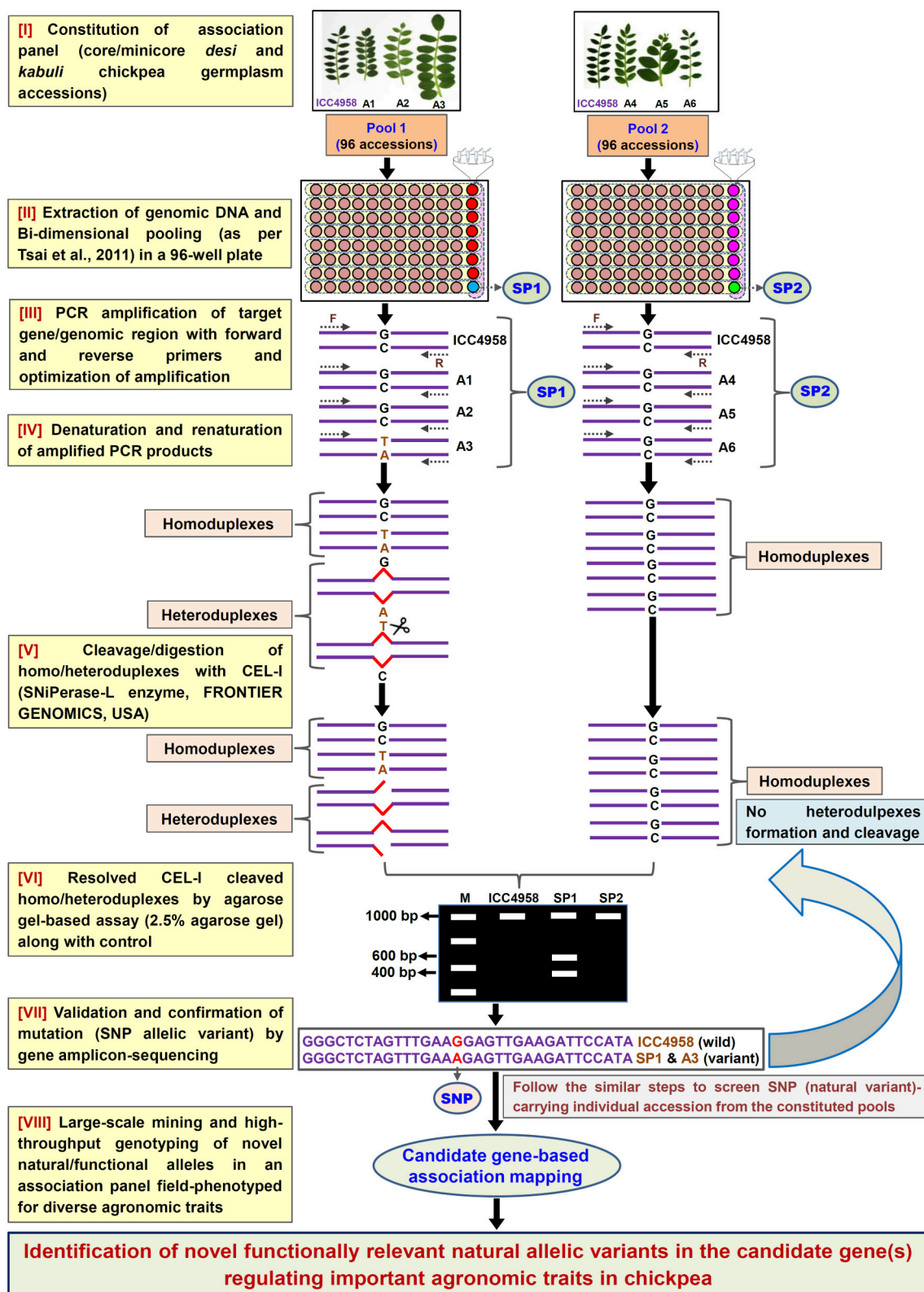


FIGURE 1 | Schematic depicting the major steps followed in an agarose gel-based EcoTILLING assay for efficient identification of functionally relevant molecular tags governing useful agronomic traits in chickpea. This strategy is optimized for successful large-scale mining of novel SNP allelic variants from the target genomic regions (genes) by genotyping in a constituted field-phenotyped association panel (*desi* and *kabuli* core/mini-core germplasm lines). A, Accessions; SP, Superpool; F, (Forward); and R, (Reverse) primers.

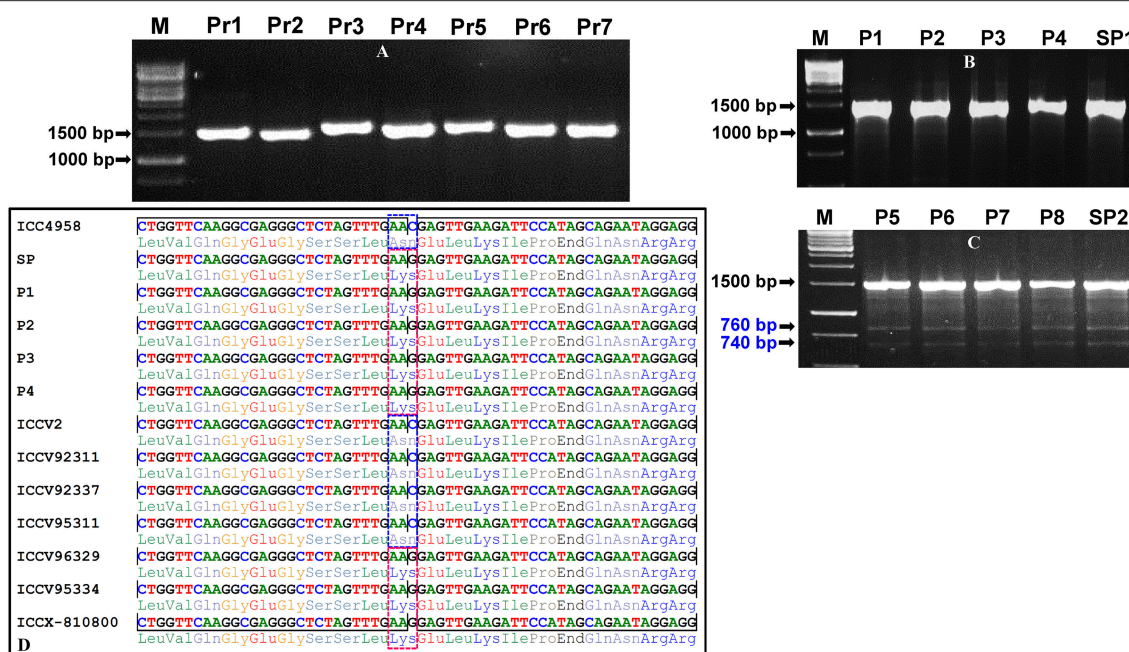


FIGURE 2 | Optimization and validation of pool-based EcoTILLING approach coupled with agarose gel detection assay for large-scale mining of novel allelic variants from diverse coding and regulatory sequence components of TF genes by genotyping in a 192 *desi* and *kabuli* germplasm lines belonging to a seed weight association panel. (A) A representative gel illustrating the optimization followed by PCR amplification of seven primer-pairs (Pr1–Pr7) designed targeting various coding and regulatory regions of seven TF genes in the genomic DNA of a *desi* chickpea accession (ICC 4958) to produce single reproducible amplicons of each primer for EcoTILLING analysis. (B,C) The representative gels depicting the screening of allelic variants from the eight representative micropools (P1–P8) and two superpools (SP1 and SP2) made from the genomic DNA of 192 *desi* and *kabuli* germplasm lines (including ICC 4958 as control) employing an agarose gel-based EcoTILLING assay as defined in the Figure 1. The absence (B) and presence (C) of one non-synonymous SNP allelic variant in the pools and superpools based on cleavage/digestion patterns of 1500 bp fragments amplified from the target CDS region of a mTERF TF gene was apparent in 2.5% agarose gel. The occurrence of 1500 bp homoduplex uncut PCR amplicons as well as mismatch-specific CEL I cleavage of 1500 bp heteroduplex PCR amplified fragments into two varied amplicons of 760 and 740 bp fragment sizes due to the effect of single nucleotide polymorphism (SNP-allele) was observed in the four pools (P5–P8) and one superpool (SP2). M: 1 kb DNA ladder size standard. (D) The sequencing of 1500 bp amplified PCR product of an mTERF TF gene followed by multiple alignment of their high-quality sequences ascertained the presence of one coding SNP (C to G) exhibiting missense non-synonymous amino acid substitution [asparagine (AAC) to lysine (AAG)] in the four pools (P5–P8), one superpool (SP2) and seven individual *desi* and *kabuli* accessions as per expectation based on agarose gel-based EcoTILLING assay. The sequenced region carrying the non-synonymous SNP is indicated with a dotted box. The detail information of genes used for validation is mentioned in the Table S2.

synonymous and missense/nonsense non-synonymous amino acid substitutions, respectively in the CDS regions of 1108 TF genes. The remaining 25 (2.2%) SNP alleles were derived from the regulatory (URR/DRR) sequence components of 25 TF genes (Table S2). To determine the physical localization (bp) of SNPs on the chickpea genome, the 100-bp TF gene sequences flanking the 1133 SNP loci were BLAST searched ($\geq 95\%$ query coverage and percent identity) against the draft genome sequences of *desi* (Jain et al., 2013) and *kabuli* (Varshney et al., 2013a) chickpea. Notably, 1042 (92%) and 91 (8%) SNPs of the total discovered 1133 TF gene-derived SNP alleles were physically mapped on the eight chromosomes and unanchored scaffolds of *desi* and *kabuli* chickpea genomes, respectively (Table S2). These observations overall infer the efficacy of agarose gel-based EcoTILLING assay in large-scale mining and high-throughput genotyping of natural as well as functional allelic variants among diverse *desi* and *kabuli* chickpea germplasm accessions by the optimal expense of time, labor and cost in the research laboratories equipped with limited infrastructural facilities. Notably, this approach

seems quite convenient and straightforward for screening the allelic variants more efficiently from the constituted pools containing DNA of numerous germplasm accessions (whole association panel) in a diploid crop species like chickpea with narrow genetic base and low intra-/inter-specific genetic polymorphism. Henceforth, this agarose-based detection assay has potential utility not only for the analysis of EcoTILLING using the pools of natural germplasm accessions but also for TILLING involving the pools of available EMS (ethyl methanesulfonate)-induced mutant lines ($\sim 10,000$) of *desi* accession (ICC 4958; Varshney et al., 2009, 2013b) to identify the functionally relevant novel SNP allelic variants (mutations) influencing vital agronomic traits. Therefore, this optimized strategy has utility in accelerating the genomics-assisted crop improvement of chickpea through genetic/association mapping. In the present study, large-scale genotyping data of novel TF gene-based SNP alleles discovered from a seed weight association panel (192 accessions) using an optimized pool-based agarose gel-EcoTILLING strategy were assessed for trait association mapping potential to identify functional and natural allelic

TABLE 1 | Eight seed weight-associated SNP allelic variants of transcription factor genes delineated by EcoTILLING-based trait association mapping.

SNP IDs	Desi and Kabuli chromosomes/scaffolds	SNP physical positions (bp)	SNPs	Gene accession IDs	Structural annotation	Putative functions	Association analysis	
							P	R ² (%)
SNP0018*	Ca_Kabuli_Chr01	12794147	[G/T]	Ca_02472	CDS-Non Syn	Basic-leucine zipper (bZIP) protein	1.3×10^{-5}	15
SNP0361*	Ca_Kabuli_Chr01	26026817	[C/T]	Ca_18591	CDS-Syn	Squamosa promoter binding protein (SBP)	2.1×10^{-4}	10
SNP0748	Ca_Desj_Scaffold_8872	3778	[C/T]	Ca_20116	URR	Zinc finger-domain containing protein	2.3×10^{-4}	10
SNP0777*	Ca_Desj_Chr01	16680465	[T/C]	Ca_TC09197	CDS-Non Syn	No apical meristem Arabidopsis transcription activation factor-cup shaped cotyledon (NAC)	1.7×10^{-4}	12
SNP0790	Ca_Desj_Chr01	46718906	[C/G]	Ca_TC09680	CDS-Non Syn	mitochondrial transcription termination factor (mTERF)	1.9×10^{-4}	11
SNP0804*	Ca_Desj_Chr02	10330226	[T/C]	Ca_TC11796	CDS-Non Syn	Basic helix-loop-helix (bHLH) protein	1.0×10^{-5}	14
SNP0875	Ca_Desj_Chr03	39546588	[G/A]	Ca_TC17366	CDS-Non Syn	APETALA-2/ethylene response element binding protein (AP2-EREBP)	1.5×10^{-4}	13
SNP0952	Ca_Desj_Chr04	48384848	[C/T]	Ca_TC03708	CDS-Non Syn	ARF (auxin response factor)	2.5×10^{-4}	10

CDS, coding DNA sequence; Syn, synonymous; and NonSyn, non-synonymous.

*Validated previously by seed weight QTL mapping.

variants of the candidate TF genes regulating seed weight in chickpea.

ECOTILLING-BASED ASSOCIATION MAPPING DELINEATES NATURALLY OCCURRING FUNCTIONAL ALLELIC VARIANTS OF CANDIDATE GENES REGULATING QUANTITATIVE TRAITS IN CHICKPEA

To perform candidate gene-based association analysis, the genotyping information of 1133 TF gene-derived SNP alleles ($\geq 5\%$ minor allele frequency) mined by EcoTILLING was integrated with multi-location replicated SW field phenotyping (100 seed weight: 6–63 g), principal component analysis (P), population genetic structure (Q), and kinship (K) matrix of 192 *desi* and *kabuli* accessions (association panel) of chickpea. At the most, we could expect clustering of 192 accessions into two distinct population groups at $K = 2$, in accordance with our preliminary genetic distance-based phylogenetic tree analysis. Using population genetic structure, the average likelihood value [$\ln P(D)$] against each K across 20 independent replications was estimated and plotted. The optimal value of K was determined following *ad hoc* and *delta K* procedures of Pritchard et al. (2000) and Evanno et al. (2005), respectively. At the optimum

value of $K = 2$, the population structure model representing expected phylogenetic relationships among 192 accessions was constructed. The principal component analysis (PCA) among accessions was performed using GAPIT (Lipka et al., 2012). The kinship matrix (K) was estimated using SPAGeDi 1.2 (Hardy and Vekemans, 2002). For candidate gene-based association analysis, the CMLM (compressed mixed linear model) ($P + K$, K and $Q + K$) along with P3D (population parameters previously determined, Kang et al., 2010; Zhang et al., 2010) interfaces of GAPIT were employed following Kujur et al. (2013, 2014), Thudi et al. (2014) and Kumar et al. (2015). To ensure the accuracy and robustness of each SNP marker-trait association, the quantile-quantile plot-based false discovery rate (FDR cut-off ≤ 0.05) corrections (Benjamini and Hochberg, 1995) for multiple comparisons between observed/expected $-\log_{10}(P)$ -values and adjusted P -value threshold of significance were performed in accordance with Kujur et al. (2015). The degree of association of SNP loci with SW trait was measured by the R^2 (model with the SNP and adjusted P -value following FDR-controlling method). The TF gene-derived SNP loci exhibiting significant association with SW trait at lowest FDR adjusted P -values (threshold $P < 10^{-4}$) and highest R^2 were identified in chickpea.

The CMLM and P3D/EMMAX-based association analysis at a FDR cut-off ≤ 0.05 detected eight TF gene-derived SNPs exhibiting significant association with 100-seed weight

at a $P \leq 10^{-4}$ (Table 1, Figure S1). Seven and one of these eight SW-associated SNPs were derived from the diverse coding (six non-synonymous and one synonymous SNP loci) and regulatory (URR) sequence components of eight TF genes, respectively. Seven SW-associated TF gene-based SNPs were physically mapped on four *desi* and *kabuli* chickpea chromosomes (1, 2, 3, and 4), whereas one SNP mapped on the unanchored scaffold of *desi* genome (Table 1, Figure S1). The proportion of SW phenotypic variation explained by eight SNP loci derived from eight TF genes [encoding *bZIP* (Basic-leucine zipper), *SBP* (Squamosa promoter binding protein) protein, Zinc finger-domain containing protein, NAC (No apical meristem arabidopsis transcription activation factor-cup shaped cotyledon), *bHLH* (Basic helix-loop-helix) protein, *AP2-EREBP* (APETALA-2/ethylene response element binding protein), *ARF* (auxin response factor), and *mTERF* (mitochondrial transcription termination factor)] among 192 *desi* and *kabuli* accessions belonging to an association panel varied from R^2 : 10 to 15% (Table 1, Figure S1). All significant eight SNP loci in combination explained 31% SW phenotypic variation. Strong association of one non-synonymous SNP in a *bZIP* TF gene (R^2 : 15% with P : 1.3×10^{-5}) with SW was observed in *desi* and *kabuli* chickpea (Table 1, Figure S1). The SW-associated eight TF genes delineated by EcoTILLING-based allele mining, genotyping and trait association mapping in chickpea probably regulate seed growth and development, including determination of seed size/weight in many crop plants (Manning et al., 2006; Agarwal et al., 2007, 2011; Nijhawan et al., 2008; Libault et al., 2009; Wang et al., 2011, 2015; Heang and Sassa, 2012; Martínez-Andújar et al., 2012; Jones and Vodkin, 2013; Ha et al., 2014; Hudson and Hudson, 2015; Liu et al., 2015; Singh and Jain, 2015; Zhang et al., 2015). Especially, the seed weight trait association potential of four TFs (*bZIP*, *SBP*, *NAC*, and *bHLH*)-derived SNPs mapped on chromosomes 1 and 2, has been ascertained by recent studies in chickpea through identification of similar gene models-containing TFs, integrating seed weight trait-specific association analysis with QTL mapping, differential expression profiling and LD (linkage disequilibrium)-based marker haplotyping (Kujur et al., 2013,

2014). The validation of these TF gene-based SNPs in two of our independent studies suggests the potential significance and robustness of these identified novel functional molecular tags (natural allelic variants and genes) in controlling seed weight, which can essentially be deployed for marker-assisted genetic enhancement of chickpea.

Collectively, the present study demonstrated the efficacy of an optimized pool-based agarose gel-EcoTILLING strategy (Figure 1) for high-throughput allele mining and genotyping as well as trait association analysis in a natural association panel to delineate novel functional allelic variants of the TF genes governing seed weight in chickpea. Therefore, this approach has potential utility to expedite various genomics-assisted breeding applications, including genetic enhancement targeting diverse qualitative and quantitative stress tolerance and yield component traits by optimal resource expenses in chickpea.

AUTHOR CONTRIBUTIONS

DB, RS, and MN conducted experiments and drafted the manuscript. ST, CB, and HU helped in constitution of association panel and performed phenotyping. SP and AT conceived and designed the study, guided data analysis and interpretation, participated in drafting and correcting the manuscript critically and gave the final approval of the version to be published. All authors have read and approved the final manuscript.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support by the Department of Biotechnology (DBT), Government of India, through their research grant (102/IFD/SAN/2161/2013-14) for this research work.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2016.00450>

REFERENCES

- Abbo, S., Berger, J., and Turner, N. C. (2003). Evolution of cultivated chickpea: four bottlenecks limit diversity and constrain adaptation. *Funct. Plant Biol.* 30, 1081–1087. doi: 10.1071/FP03084
- Abbo, S., Gopher, A., Rubin, B., and Lev-Yadun, S. (2005). On the origin of Near Eastern founder crops and the 'dump-heap hypothesis'. *Genet. Res. Crop Evol.* 52, 491–495. doi: 10.1007/s10722-004-7069-x
- Agarwal, P., Arora, R., Ray, S., Singh, A. K., Singh, V. P., Takatsuji, H., et al. (2007). Genome-wide identification of C2H2 zinc-finger gene family in rice and their phylogeny and expression analysis. *Plant Mol. Biol.* 65, 467–485. doi: 10.1007/s11103-007-9199-y
- Agarwal, P., Kapoor, S., and Tyagi, A. K. (2011). Transcription factors regulating the progression of monocot and dicot seed development. *Bioessays* 33, 189–202. doi: 10.1002/bies.201000107
- Bajaj, D., Saxena, M. S., Kujur, A., Das, S., Badoni, S., Tripathi, S., et al. (2015). Genome-wide conserved non-coding microsatellite (CNMS) marker-based integrative genetical genomics for quantitative dissection of seed weight in chickpea. *J. Exp. Bot.* 66, 1271–1290. doi: 10.1093/jxb/eru478
- Barkley, N. A., and Wang, M. L. (2008). Application of TILLING and EcoTILLING as reverse genetic approaches to elucidate the function of genes in plants and animals. *Curr. Genomics* 9, 212–226. doi: 10.2174/138920208784533656
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. Ser. B* 57, 289–300.
- Berger, J. D., Abbo, S., and Turner, N. C. (2003). Ecogeography of annual wild *Cicer* species: the poor state of the world collection. *Crop Sci.* 43, 1076–1090. doi: 10.2135/cropsci2003.1076
- Berger, J. D., Buck, R., Henzell, J. M., and Turner, N. C. (2005). Evolution in the genus *Cicer* vernalisation response and low temperature pod set in chickpea (*C. arietinum* L.) and its annual wild relatives. *Aust. J. Agric. Res.* 56, 1191–1200. doi: 10.1071/AR05089
- Caldwell, D. G., McCallum, N., Shaw, P., Muehlbauer, G. J., Marshall, D. F., and Waugh, R. (2004). A structured mutant population for forward and

- reverse genetics in barley (*Hordeum vulgare* L.). *Plant J.* 40, 143–150. doi: 10.1111/j.1365-3113.2004.02190.x
- Comai, L., Young, K., Till, B. J., Reynolds, S. H., Greene, E. A., Codomo, C. A., et al. (2004). Efficient discovery of DNA polymorphisms in natural populations by EcoTILLING. *Plant J.* 37, 778–786. doi: 10.1111/j.0960-7412.2003.01999.x
- Evanno, G., Regnaut, S., and Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.* 14, 2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Frerichmann, S. L., Kirchhoff, M., Müller, A. E., Scheidig, A. J., Jung, C., and Kopisch-Obuch, F. J. (2013). EcoTILLING in *Beta vulgaris* reveals polymorphisms in the FLC-like gene *BvFL1* that are associated with annuality and winter hardiness. *BMC Plant Biol.* 13:52. doi: 10.1186/1471-2229-13-52
- Ha, C. V., Esfahani, M. N., Watanabe, Y., Tran, U. T., Sulieman, S., Mochida, K., et al. (2014). Genome-wide identification and expression analysis of the CaNAC family members in chickpea during development, dehydration and ABA treatments. *PLoS ONE* 9:e114107. doi: 10.1371/journal.pone.0114107
- Hardy, O. J., and Vekemans, X. (2002). SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* 2, 618–620. doi: 10.1046/j.1471-8286.2002.00305.x
- Heang, D., and Sassa, H. (2012). Antagonistic actions of *HLH/bHLH* proteins are involved in grain length and weight in rice. *PLoS ONE* 7:e31325. doi: 10.1371/journal.pone.0031325
- Henikoff, S., Till, B. J., and Comai, L. (2004). TILLING: traditional mutagenesis meets functional genomics. *Plant Physiol.* 135, 630–636. doi: 10.1104/pp.104.041061
- Hudson, K. A., and Hudson, M. E. (2015). A classification of basic helix-loop-helix transcription factors of soybean. *Int. J. Genomics* 2015:603182. doi: 10.1155/2015/603182
- Ibiza, V. P., Cañizares, J., and Nuez, F. (2010). EcoTILLING in *Capsicum* species: searching for new virus resistances. *BMC Genomics* 11:631. doi: 10.1186/1471-2164-11-631
- Jain, M., Misra, G., Patel, R. K., Priya, P., Jhanwar, S., Khan, A. W., et al. (2013). A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). *Plant J.* 74, 715–729. doi: 10.1111/tj.12173
- Jhanwar, S., Priya, P., Garg, R., Parida, S. K., Tyagi, A. K., and Jain, M. (2012). Transcriptome sequencing of wild chickpea as a rich resource for marker development. *Plant Biotechnol. J.* 10, 690–702. doi: 10.1111/j.1467-7652.2012.00712.x
- Jones, S. I., and Vodkin, L. O. (2013). Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS ONE* 8:e59270. doi: 10.1371/journal.pone.0059270
- Kang, H. M., Sul, J. H., Service, S. K., Zaitlen, N. A., Kong, S. Y., Freimer, N. B., et al. (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat. Genet.* 42, 348–354. doi: 10.1038/ng.548
- Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2013). Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA Res.* 20, 355–374. doi: 10.1093/dnares/dst015
- Kujur, A., Bajaj, D., Saxena, M. S., Tripathi, S., Upadhyaya, H. D., Gowda, C. L. L., et al. (2014). An efficient and cost-effective approach for genic microsatellite marker-based large-scale trait association mapping: identification of candidate genes for seed weight in chickpea. *Mol. Breed.* 34, 241–265. doi: 10.1007/s11032-014-0033-3
- Kujur, A., Bajaj, D., Upadhyaya, H. D., Das, S., Ranjan, R., Shree, T., et al. (2015). Employing genome-wide SNP discovery and genotyping strategy to extrapolate the natural allelic diversity and domestication patterns in chickpea. *Front. Plant Sci.* 6:162. doi: 10.3389/fpls.2015.00162
- Kumar, V., Singh, A., Amitha Mithra, S. V., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22, 133–145. doi: 10.1093/dnares/dsu046
- Libault, M., Joshi, T., Benedito, V. A., Xu, D., Udvardi, M. K., and Stacey, G. (2009). Legume transcription factor genes: what makes legumes so special? *Plant Physiol.* 151, 991–1001. doi: 10.1104/pp.109.144105
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28, 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, K., Yuan, C., Li, H., Lin, W., Yang, Y., Shen, C., et al. (2015). Genome-wide identification and characterization of auxin response factor (ARF) family genes related to flower and fruit development in papaya (*Carica papaya* L.). *BMC Genomics* 16:901. doi: 10.1186/s12864-015-2182-0
- Manning, K., Tör, M., Poole, M., Hong, Y., Thompson, A. J., King, G. J., et al. (2006). A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat. Genet.* 38, 948–952. doi: 10.1038/ng1841
- Martínez-Andújar, C., Martín, R. C., and Nonogaki, H. (2012). Seed traits and genes important for translational biology-highlights from recent discoveries. *Plant Cell Physiol.* 53, 5–15. doi: 10.1093/pcp/pcr112
- McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeted screening for induced mutations. *Nat. Biotechnol.* 18, 455–457. doi: 10.1038/74542
- Mejlhede, N., Kyjovska, Z., Backes, G., Burhenne, K., Rasmussen, S. K., and Jahoor, A. (2006). EcoTILLING for the identification of allelic variation in the powdery mildew resistance genes *mlo* and *Mla* of barley. *Plant Breed.* 125, 461–467. doi: 10.1111/j.1439-0523.2006.01226.x
- Negrao, S., Courtois, B., Ahmadi, N., Abreu, I., Saibo, N., and Oliveira, M. M. (2011). Recent updates on salinity stress in rice: from physiological to molecular responses. *Crit. Rev. Plant Sci.* 30, 329–377. doi: 10.1080/07352689.2011.587725
- Nijhawan, A., Jain, M., Tyagi, A. K., and Khurana, J. P. (2008). Genomic survey and gene expression analysis of the basic leucine zipper transcription factor family in rice. *Plant Physiol.* 146, 333–350. doi: 10.1104/pp.107.112821
- Perry, J. A., Wang, T. L., Welham, T. J., Gardner, S., Pike, J. M., Yoshida, S., et al. (2003). A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiol.* 131, 866–871. doi: 10.1104/pp.102.017384
- Pritchard, J. K., Stephens, M., and Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* 155, 945–959.
- Raghavan, C., Naredo, M. E. B., Wang, H., Atienza, G., Liu, B., Qiu, F., et al. (2007). Rapid method for detecting SNPs on agarose gels and its application in candidate gene mapping. *Mol. Breed.* 19, 87–101. doi: 10.1007/s11032-006-9046-x
- Saxena, M. S., Bajaj, D., Das, S., Kujur, A., Kumar, V., Singh, M., et al. (2014b). An integrated genomic approach for rapid delineation of candidate genes regulating agro-morphological traits in chickpea. *DNA Res.* 21, 695–710. doi: 10.1093/dnares/dsu031
- Saxena, M. S., Bajaj, D., Kujur, A., Das, S., Badoni, S., Kumar, V., et al. (2014a). Natural allelic diversity, genetic structure and linkage disequilibrium pattern in wild chickpea. *PLoS ONE* 9:e107484. doi: 10.1371/journal.pone.0107484
- Singh, R., Sharma, P., Varshney, R. K., and Sharma, S. K. (2008). Chickpea improvement: role of wild species and genetic markers. *Biotechnol. Genet. Eng.* 25, 267–313. doi: 10.5661/bger-25-267
- Singh, V. K., and Jain, M. (2015). Genome-wide survey and comprehensive expression profiling of Aux/IAA gene family in chickpea and soybean. *Front. Plant Sci.* 6:918. doi: 10.3389/fpls.2015.00918
- Suzuki, T., Eiguchi, M., Satoh, H., Kumamaru, T., and Kurata, N. (2005). A modified TILLING system for rice mutant screening. *Rice Genet. News Lett.* 22, 89–91.
- Thudi, M., Upadhyaya, H. D., Rathore, A., Gaur, P. M., Krishnamurthy, L., Roorkiwal, M., et al. (2014). Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping approaches. *PLoS ONE* 9:e96758. doi: 10.1371/journal.pone.0096758
- Till, B. J., Cooper, J., Tai, T. H., Colowit, P., Greene, E. A., Henikoff, S., et al. (2007). Discovery of chemically induced mutations in rice by TILLING. *BMC Plant Biol.* 7:19. doi: 10.1186/1471-2229-7-19
- Till, B. J., Jankowicz-Cieslak, J., Sagi, L., Huynh, O. A., Utsushi, H., Swennen, R., et al. (2010). Discovery of nucleotide polymorphisms in the *Musa* gene pool by EcoTILLING. *Theor. Appl. Genet.* 121, 1381–1389. doi: 10.1007/s00122-010-1395-5
- Till, B. J., Zerr, T., Comai, L., and Henikoff, S. (2006). A protocol for TILLING and EcoTILLING in plants and animals. *Nat. Protoc.* 1, 2465–2477. doi: 10.1038/nprot.2006.329
- Toker, C. (2009). A note on the evolution of *kabuli* chickpeas as shown by induced mutations in *Cicer reticulatum* Ladizinsky. *Genet. Resour. Crop. Evol.* 56, 7–12. doi: 10.1007/s10722-008-9336-8

- Tsai, H., Howell, T., Nitcher, R., Missirian, V., Watson, B., Ngo, K. J., et al. (2011). Discovery of rare mutations in populations: TILLING by sequencing. *Plant Physiol.* 156, 1257–1268. doi: 10.1104/pp.110.169748
- Upadhyaya, H. D., Bramel, P. J., and Singh, S. (2001). Development of a chickpea core subset using geographic distribution and quantitative traits. *Crop Sci.* 41, 206–210. doi: 10.2135/cropsci2001.411206x
- Upadhyaya, H. D., Dwivedi, S. L., Baum, M., Varshney, R. K., Udupa, S. M., Gowda, C. L. L., et al. (2008). Genetic structure, diversity, and allelic richness in composite collection and reference set in chickpea (*Cicer arietinum* L.). *BMC Plant Biol.* 8:106. doi: 10.1186/1471-2229-8-106
- Upadhyaya, H. D., and Ortiz, R. (2001). A mini-core subset for capturing diversity and promoting utilization of chickpea genetic resources in crop improvement. *Theor. Appl. Genet.* 102, 1292–1298. doi: 10.1007/s00122-001-0556-y
- Upadhyaya, H. D., Ortiz, R., Bramel, P. J., and Singh, S. (2002). Phenotypic diversity for morphological and agronomic characteristics in chickpea core collection. *Euphytica* 123, 333–342. doi: 10.1023/A:1015088417487
- Upadhyaya, H. D., Thudi, M., Dronavalli, N., Gujaria, N., Singh, S., Sharma, S., et al. (2011). Genomic tools and germplasm diversity for chickpea improvement. *Plant Genetic Res.* 9, 45–58. doi: 10.1017/S1479262110000468
- Varshney, R. K., Close, T. J., Singh, N. K., Hoisington, D. A., and Cook, D. R. (2009). Orphan legume crops enter the genomics era! *Curr. Opin. Plant Biol.* 12, 202–210. doi: 10.1016/j.pbi.2008.12.004
- Varshney, R. K., Mohan, S. M., Gaur, P. M., Gangarao, N. V., Pandey, M. K., and Bohra, A. (2013b). Achievements and prospects of genomics-assisted breeding in three legume crops of the semi-arid tropics. *Biotechnol. Adv.* 31, 1120–1134. doi: 10.1016/j.biotechadv.2013.01.001
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013a). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Wang, J. W., Park, M. Y., Wang, L. J., Koo, Y., Chen, X. Y., Weigel, D., et al. (2011). miRNA control of vegetative phase change in trees. *PLoS Genet.* 7:e1002012. doi: 10.1371/journal.pgen.1002012
- Wang, N., Shi, L., Tian, F., Ning, H., Wu, X., Long, Y., et al. (2010). Assessment of *FAE1* polymorphisms in three *Brassica* species using EcoTILLING and their association with differences in seed erucic acid contents. *BMC Plant Biol.* 10:137. doi: 10.1186/1471-2229-10-137
- Wang, Z., Cheng, K., Wan, L., Yan, L., Jiang, H., Liu, S., et al. (2015). Genome-wide analysis of the basic leucine zipper (bZIP) transcription factor gene family in six legume genomes. *BMC Genomics* 16:1053. doi: 10.1186/s12864-015-2258-x
- Xia, Y., Ning, Z., Bai, G., Li, R., Yan, G., Siddique, K. H., et al. (2012). Allelic variations of a light harvesting chlorophyll a/b-binding protein gene (*Lhcb1*) associated with agronomic traits in barley. *PLoS ONE* 7:e37573. doi: 10.1371/journal.pone.0037573
- Yang, W., Bai, X., Kabelka, E., Eaton, C., Kamoun, S., van der Knaap, E., et al. (2004). Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. *Mol. Breed.* 14, 21–34. doi: 10.1023/B:MOLB.0000037992.03731.a5
- Yu, S., Liao, F., Wang, F., Wen, W., Li, J., Mei, H., et al. (2012). Identification of rice transcription factors associated with drought tolerance using the EcoTILLING method. *PLoS ONE* 7:e30765. doi: 10.1371/journal.pone.0030765
- Zhang, S. D., Ling, L. Z., and Yi, T. S. (2015). Evolution and divergence of SBP-box genes in land plants. *BMC Genomics* 16:787. doi: 10.1186/s12864-015-1998-y
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Bajaj, Srivastava, Nath, Tripathi, Bharadwaj, Upadhyaya, Tyagi and Parida. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Gene Classification and Mining of Molecular Markers Useful in Red Clover (*Trifolium pratense*) Breeding

Jan Ištváněk¹, Jana Dluhošová¹, Petr Dluhoš², Lenka Pátková¹, Jan Nedělník³ and Jana Řepková^{1*}

¹ Department of Experimental Biology, Faculty of Science, Masaryk University, Brno, Czechia, ² Department of Psychiatry, University Hospital Brno and Masaryk University, Brno, Czechia, ³ Agricultural Research, Ltd., Troubsko, Czechia

OPEN ACCESS

Edited by:

Diego Rubiales,
Instituto de Agricultura Sostenible
(CSIC), Spain

Reviewed by:

Leif Skot,
Aberystwyth University, UK
R. Varma Penmetsa,
University of California at Davis, USA

*Correspondence:

Jana Řepková
repkova@sci.muni.cz

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 19 November 2016

Accepted: 01 March 2017

Published: 22 March 2017

Citation:

Ištváněk J, Dluhošová J, Dluhoš P,
Pátková L, Nedělník J and Řepková J
(2017) Gene Classification and Mining
of Molecular Markers Useful in Red
Clover (*Trifolium pratense*) Breeding.
Front. Plant Sci. 8:367.
doi: 10.3389/fpls.2017.00367

Red clover (*Trifolium pratense*) is an important forage plant worldwide. This study was directed to broadening current knowledge of red clover's coding regions and enhancing its utilization in practice by specific reanalysis of previously published assembly. A total of 42,996 genes were characterized using Illumina paired-end sequencing after manual revision of Blast2GO annotation. Genes were classified into metabolic and biosynthetic pathways in response to biological processes, with 7,517 genes being assigned to specific pathways. Moreover, 17,727 enzymatic nodes in all pathways were described. We identified 6,749 potential microsatellite loci in red clover coding sequences, and we characterized 4,005 potential simple sequence repeat (SSR) markers as generating polymerase chain reaction products preferentially within 100–350 bp. Marker density of 1 SSR marker per 12.39 kbp was achieved. Aligning reads against predicted coding sequences resulted in the identification of 343,027 single nucleotide polymorphism (SNP) markers, providing marker density of one SNP marker per 144.6 bp. Altogether, 95 SSRs in coding sequences were analyzed for 50 red clover varieties and a collection of 22 highly polymorphic SSRs with pooled polymorphism information content >0.9 was generated, thus obtaining primer pairs for application to diversity studies in *T. pratense*. A set of 8,623 genome-wide distributed SNPs was developed and used for polymorphism evaluation in individual plants. The polymorphic information content ranged from 0 to 0.375. Temperature switch PCR was successfully used in single-marker SNP genotyping for targeted coding sequences and for heterozygosity or homozygosity confirmation in validated five loci. Predicted large sets of SSRs and SNPs throughout the genome are key to rapidly implementing genome-based breeding approaches, for identifying genes underlying key traits, and for genome-wide association studies. Detailed knowledge of genetic relationships among breeding material can also be useful for breeders in planning crosses or for plant variety protection. Single-marker assays are useful for diagnostic applications.

Keywords: biosynthetic pathways, genetic diversity, sequencing, SNP, specific genes, SSR

INTRODUCTION

Fabaceae is among the most studied of plant families. The third-largest plant family, it includes many food and industrial plants and stands second only to Poaceae among the most important plant families from economic and nutritional perspectives (Graham and Vance, 2003). This importance results not only from the species' economic and nutritive values, but also from their unique capability for fixing atmospheric nitrogen. In the past decade, the extent of genomic information available on legumes has been broadened substantially. Such model species as *Medicago truncatula* Gaertn. (Young et al., 2011) and *Lotus japonicus* L. (Sato et al., 2008), as well as the crops soybean (*Glycine max* [L.] Merrill.; Schmutz et al., 2010), pigeon pea (*Cajanus cajan* [L.] Millsp.; Varshney et al., 2012), and chickpea (*Cicer arietinum* L.; Varshney et al., 2013) have been sequenced. Several other sequencing projects are under way which encompass a broad range of agronomically and horticulturally important plants (www.phytozome.net).

Red clover (*Trifolium pratense* L.) belongs to the tribe Trifolieae, together with another 240 annual and perennial herb species, both wild and cultivated. It is an important forage plant worldwide, serving as a temporary cover crop or manure crop as well as for silage production and grazing. Like other legumes, it is capable of fixing atmospheric nitrogen via symbiosis with *Rhizobium leguminosarum* bv. *trifolii* (Sprent, 2009). Its breeding and related research have been complicated, however, by the species' outcrossing nature with gametophytic self-incompatibility. The resulting heterozygosity has hampered intensive genetic and genomic analysis. Nonetheless, with the rising availability of sequencing technology, red clover has been a target of several genomic studies in recent years.

Red clover's nuclear genome is divided into seven chromosomes ($x = 7$) with size estimated to be 418 Mbp ($1C = 0.43$ pg; Vižintin et al., 2006). The first consensus high-density linkage map contained 1,414 simple sequence repeats (SSRs), 181 amplified fragment length polymorphisms, and 228 restriction fragment length polymorphisms (Isobe et al., 2009). The structure of the red clover genome has been investigated using fluorescence *in situ* hybridization (Sato et al., 2005; Kataoka et al., 2012). The genome also has been compared with those of related species (white clover, *M. truncatula* and *L. japonicus*) using DNA markers (Isobe et al., 2012). DNA markers, too, can be used in various research and practical approaches. For example, two studies used DNA markers to identify quantitative trait loci (QTLs) related to persistence (Herrmann et al., 2008), disease resistance, and winter hardiness (Klimenko et al., 2010) in full sib mapping families. Recently, great insight into red clover genomics has been achieved through application of next-generation sequencing (NGS) technology. Both whole-genome sequencing (WGS; Ištvánek et al., 2014; De Vega et al., 2015) and RNA sequencing (Yates et al., 2014) have been carried out in red clover. While WGS focused on describing red clover's genome, RNA sequencing described transcriptome differences in conditions of drought stress. Concurrently, studies of both types identified a great number of DNA markers which can be of great value in practical applications.

As a consequence of the outcrossing, both natural ecotypes and varieties that may be morphologically similar are likely to be highly heterogeneous genetically. Strategies for genetic diversity analysis based on DNA profiling must address this issue and enable quantification of variation within and among populations. Evaluation of genetic variation for outcrossing forage species is important for the processes of cultivar identification and seed purity analysis, ecological analysis of pasture populations, and selection of genetically divergent parents for genetic mapping studies (Forster et al., 2001). The genetic divergence of some genotypes ensures a high level of genetic polymorphism in crosses. Breeding methods for cross-pollinated forage crops, including red clover, require strategies for genotyping. Genetic markers assaying variation in transcribed regions of genes with known functions will be useful for developing trait-linked markers. NGS has shown great potential for large-scale production of functional genes and molecular markers at the whole-genome level, especially in non-model organisms. Two important tasks for NGS are identifying expression patterns in biochemical processes and classifying genes into specific pathways. Legumes can produce more secondary metabolites (especially cyanogenic glucosides, glucosinolates, amines, and alkaloids) than can other plants which are not nitrogen fixers. Most secondary metabolites exhibit some biological, pharmacological, or toxicological activity (Teuscher and Lindequist, 2010; Wink, 2013). In this respect, the Fabaceae are distinguished by isoflavones, which function as antioxidants, phytoestrogens, and antimicrobial compounds. The benefits of protecting plant proteins from degradation in the rumen by means of polyphenol oxidases (PPOs) have been established in some fodder crops, and red clover contains PPOs in significant quantities (Jones et al., 1995; Jakešová et al., 2015). There is also an increasing need to develop molecular markers for resistance genes or components relating to nitrogen fixation.

Based on previously published genome assembly (Ištvánek et al., 2014), this study aims to elucidate red clover genes involved within complex biosynthetic pathways in response to biological processes. Special attention is given to specific secondary metabolites inasmuch as they can significantly influence the final variety's breeding strategy and purpose. Gene-specific SSR and single nucleotide polymorphism (SNP) markers are reported and described with a view to enhancing marker-assisted breeding in outcrossing species of red clover. Finally, we developed and validated sets of polymorphic microsatellites and SNPs for the analysis of genetic relationships among red clover varieties and individuals. The findings of this study can be useful in investigating genetic diversity and red clover breeding. These markers will contribute to enriching the current reference red clover map, generating more informative genetic and genomic tools, and enabling genome synteny analysis.

MATERIALS AND METHODS

Sequencing and Gene Annotation

Sequencing, *de novo* assembly, gene prediction, and initial annotation is described in Ištvánek et al. (2014). Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank

under the accession ASHM00000000. The version used in this paper is ASHM01000000. Detailed inspection of gene annotation was completed manually. Protein-encoding genes were classified into functional categories according to Gene Ontology (GO) annotation, and the results were summarized in plant GOSlim functional categories. Each gene was aligned against KEGG (release 67.1) proteins, and the pathway in which that gene might be involved was determined. Genes were also sorted by the pathways.

Comparison of *Trifolium* Genes with Model Legumes

Homologous gene sequences were analyzed among red clover, *M. truncatula*, and *G. max*. Predicted genes in these species were mapped and searched for homology on the basis of red clover (DDBJ/EMBL/NCBI accession numbers: LT555306.1–LT555312.1; De Vega et al., 2015), *M. truncatula*, and *G. max* chromosome sequences. The best hits from a TBLASTX ($e \leq 1e^{-15}$) search with at least 70% identity were mapped on these chromosomes and gene densities for every 100 kbp were counted. Homologous sequences for these species were determined using the best hits from a reciprocal TBLASTX search. Circos software (Krzywinski et al., 2009) was used to visualize the top half hits of the data through a circular concentric ideograms layout. Gene densities were displayed as histogram plots and homologous sequences for the three species as lines.

DNA Marker Prediction

SSR Locator (da Maia et al., 2008) was used to mine SSRs in the red clover genes as well as for primer design. Uniform melting temperature at $T_m = 55^\circ\text{C}$ was set for all predicted SSR sites, which were defined as a monomer occurring at least 12 times, a dimer occurring at least 6 times, tri- and tetramers occurring at least 4 times, and penta- and hexamers occurring at least 3 times. The number of polymerase chain reaction (PCR) products was predicted for each primer pair.

Probable SNPs in genes were discovered by aligning Tatra reads onto predicted genes using bwa v0.7.5 (Li and Durbin, 2010). Samtools v0.1.19 (Li et al., 2009) and Picard v1.80 (<http://broadinstitute.github.io/picard/>) were used in subsequent steps of marking PCR duplicates, sorting, and indexing. GATK v2.7 (<https://www.broadinstitute.org/gatk/>) was used to remap the reads near the InDels, recalibrate base quality scores, and identify SNPs. Only sites with SNP calling quality scores of 30 or higher and with read depths of at least 10 reads were marked as high-quality SNPs. For subsequent filtration, custom perl scripts were used.

DNA Marker Validation

SSR marker validation was carried out for 50 red clover varieties, 34 of which were Czech varieties and 16 from other countries (Table 1). SNP marker validation was performed for 5 varieties (Amos, Fresko, Start, Tatra, Tempus). Leaves were collected from plants 30 days old which had been grown in a greenhouse. For SSR analysis, genomic DNA was isolated from 16 pooled plants per variety from ~1 g of young leaves using the modified protocol by Dellaporta et al. (1983). Concentration and purity of extracted

genomic DNA was assessed using NanoDrop (Thermo Scientific, Waltham, MA, USA).

To validate SSRs, we randomly selected 96 SSR loci in coding sequences. SSR primers (Table S1) were predicted by SSR Locator. Validations of predicted SSRs were performed via PCR and electrophoretic separation. PCRs were carried out in a volume of 10 μl with 1x reaction buffer, 0.2 mM of each dNTP (10 mM; Sigma-Aldrich, Steinheim, Germany), 10 pmol of each primer, 0.5 U of GoTaq[®] polymerase (Promega, Madison, WI, USA), and 30 ng of genomic DNA template.

Cycling conditions were set as follows: a preliminary step at 94°C for 3 min, 58°C for 1 min, 72°C for 1 min; 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s; and an elongation step at 72°C for 5 min. PCR-amplified fragments were separated by electrophoresis on either a 3% agarose gel or a 10% polyacrylamide gel and visualized by ethidium bromide staining.

The validation of predicted SNP variants was performed by SNP array (Arrayit Corporation, CA, USA) with 8,623 genome-wide distributed SNPs. We examined intra-variety genetic heterogeneity using SNP genotyping for a set of 20 DNA samples of individual red clover plants of the variety Tatra. Probes, 15-mer oligonucleotides, were designed with the SNP at the center position without overlap (two probes per SNP; Table S2). The fluorescent dyes Alexa Fluor[®] 555 and Alexa Fluor[®] 647 (Invitrogen, CA, USA) were used for labeling. Capture agents were printed into 1–48 microarrays per 25×76 mm glass substrate slide, each probe 3 times. Hybridization was performed with genomic DNA isolated from individual plants using the modified protocol by Dellaporta et al. (1983). Variance stabilizing normalization was used to evaluate fluorescence intensities of the reference and alternative alleles by RStudio software (RStudio Team, 2015) with limma package (Ritchie et al., 2015) and the log2 intensity data were processed.

Single-marker SNP polymorphisms were also validated by the modified technique of temperature switch PCR (Tabone et al., 2009) whereby two PCRs were carried out for each locus, one for a reference (R) allele and the other for an alternative (A) allele. Due to this technique's specific requirements, only SNP sites having no other SNP in their vicinity for 30 bp were included within the validation. Genomic DNA was isolated from 100 mg of leaves of individual plants using the CTAB method (Rogers and Bendich, 1989). SNP validation was performed for five candidate SNPs from Tatra coding sequences used as a reference. Primers (Table S3) for SNPs were predicted using Primer3 and OligoCalc. For an R allele, two primer pairs were used (LS—locus specific and NLS—nested locus specific) with different melting temperatures. PCR amplification using these four primers provides 4 amplicons, while only the amplicon emerging from the NLS primers (with NLS_F primer directly binding to the SNP position) confirms the presence of an R allele in a selected sample. For reliable detection of an A allele, we performed a simplified PCR reaction with two primers, a forward LS primer and a reverse primer (labeled as reverse primer SNP; SNP_R) which binds with its 3' end to a SNP nucleotide. The presence of a PCR product of a predicted length confirms the existence of an A allele. All expected products with their predicted lengths are listed in Table S4. The PCR components used for

TABLE 1 | List of red clover varieties and their characterization.

ECN*	Variety	Ploidy	Origin
13T0200096	Agil	2x	CZ
13T0200080	Bonus	2x	CZ
13T0200097	Brisk	2x	CZ
13T0200034	Chlumecky	2x	CZ
13T0230104	Concorde	2x	US
13T0200018	Essex broad red	2x	GB
13T0200081	Garant	2x	CZ
13T0200357	Gibridnij pozdnespelyj	2x	SU
13T0200492	Grasslands hamua	2x	NZ
13T0230094	Makimidori	2x	JP
13T0200625	Nemaro	2x	DE
13T0200568	Parka	2x	PL
13T0200056	Pavo	2x	CH
13T0200496	Radan	2x	CZ
13T0200127	Respect	2x	CZ
13T0200029	Slavin	2x	CZ
13T0200030	Slavoj	2x	CZ
13T0230101	Spurt	2x	CZ
13T0200319	Start	2x	CZ
13T0230006	Suez	2x	CZ
13T0200598	Tabor	2x	CZ
13T0200039	Trubadur	2x	CZ
13T0200020	Van	2x	CZ
13T0230090	Vendelin	2x	CZ
13T0200600	Vltavin	2x	CZ
13T0230095	Walter	2x	CA
13T0230023	Amos	4x	CZ
13T0200058	Astur	4x	CH
13T0200088	Atlantis	4x	DE
13T0230132	Beskyd	4x	CZ
13T0230024	Bivoj	4x	CZ
13T0230021	Blizard	4x	CZ
13T0230034	Cyklon	4x	CZ
13T0230133	Dolina	4x	CZ
13T0230039	Dolly	4x	CZ
13T0230145	Fresko	4x	CZ
13T0200571	Hungarotetra	4x	HU
13T0200329	Kvarta	4x	CZ
13T0200486	Lossam	4x	FR
13T0230026	Margot	4x	SK
13T0200476	Radegast	4x	CZ
13T0230110	Rezista	4x	CZ
13T0230120	Sigord	4x	SK
13T0230134	Sprint	4x	CZ
13T0200327	Tatra	4x	CZ
13T0200538	Tempus	4x	CZ
13T0230114	Titus	4x	CZ
13T0200460	Triton	4x	SE
13T0200597	Vesna	4x	CZ
13T0200636	Vulkan	4x	CZ

*National accession number—GeneBank of Crop Research Institute Ltd., Prague-Ruzyně, Czech Republic; CA, Canada; CH, Switzerland; CZ, Czech Republic; DE, Germany; FR, France; HU, Hungary; JP, Japan; NZ, New Zealand; PL, Poland; SE, Sweden; SK, Slovakia; SU, Soviet Union; US, United States.

both PCRs were the same as for the SSR validation, with minor modifications: for detection of the *R* allele, 5 pmol of each NLS primer and 1 pmol of each LS primer were used. Also, PCR was enriched with 1% bovine serum albumin. To detect the *A* allele, 5 pmol of both primers were used. Both PCRs used GoTaq® polymerase in a concentration of 0.25 U. Cycling conditions for the *R* allele were used according to Tabone et al. (2009). Cycling conditions for the *A* allele were shortened to a denaturation step at 95°C for 5 min; 30 cycles of 95°C for 30 s, 62°C for 30 s, and 72°C for 30 s; then 72°C for 5 min as a final elongation step. PCR-amplified fragments were separated by electrophoresis on a 10% polyacrylamide gel and visualized by ethidium bromide staining.

Polymorphism Evaluation

Evaluation of individual polymorphic fragments was inferred from agarose or polyacrylamide gels with separated PCR-amplified fragments and was performed for each SSR marker manually. Pooled polymorphism information content (pPIC) was then calculated for each SSR marker, expressing the probability of detecting a polymorphism between genotypes of two randomly drawn red clover varieties. pPIC was calculated for all 50 red clover varieties. Calculations were performed for each SSR marker *m* separately as follows:

- 1) Genotypes were divided into two groups (G_1 and G_0) according to the presence/absence of any PCR-amplified product for the marker *m* (i.e., G_1 contained all genotypes with at least one PCR-amplified product and G_0 contained the rest).
- 2) pPIC for the subgroup G_1 was calculated as:

$$pPIC_1 = 1 - \prod_{i=1}^n (1 - 2f_i(1 - f_i)),$$

where *i* is one particular band from *n* possible bands of this marker; \prod is the product operator, i.e., the symbol denoting product of a sequence in a similar manner as \sum denotes summation; and f_i is the frequency of the *i*-th band among all genotypes in G_1 . $pPIC_1$ is thus the probability that the marker *m* can distinguish any two random genotypes whose probabilities for the presence of each band come from the same distribution as in the sample G_1 .

- 3) pPIC for the subgroup G_0 was $pPIC_0 = 0$, because the genotypes in G_0 cannot be distinguished from each other (they did not have any PCR-amplified products).
- 4) pPIC for comparison of two genotypes, one from each group, was $pPIC_{1 \times 0} = 1$, because any genotype from G_1 can be distinguished from any G_0 genotype by having at least one product.
- 5) Overall pPIC for the marker *m* was then calculated by weighting the values acquired in previous steps according to the proportionate sizes of G_1 and G_0 :

$$\begin{aligned} pPIC &= p_1^2 \cdot pPIC_1 + 2p_1p_0 \cdot pPIC_{1 \times 0} + p_0^2 \cdot pPIC_0 \\ &= p_1^2 \cdot pPIC_1 + 2p_1p_0 \cdot 1 + p_0^2 \cdot 0 \\ &= p_1^2 \cdot pPIC_1 + 2p_1p_0, \end{aligned}$$

where $p_1 = \frac{N_1}{N_1 + N_0}$, $p_0 = \frac{N_0}{N_1 + N_0}$ are proportions of genotypes in the two groups G_1 and G_0 (N_1 and N_0 are counts of genotypes in G_1 and G_0 for the marker m).

The polymorphic information content (PIC) of the SNP loci was calculated according to Botstein et al. (1980).

Phylogenetic Analysis

The similarity between each pair of *T. pratense* varieties was assessed according to the presence or absence of individual separated PCR-amplified fragments using the Jaccard (1901) and Sørensen–Dice (Dice, 1945; Sørensen, 1948) indices for each SSR marker. The Jaccard and Sørensen–Dice indices were calculated as $n_{xy}/(n_x + n_y - n_{xy})$ and $2n_{xy}/(n_x + n_y)$, respectively, where n_{xy} represents the number of bands which are present simultaneously in both compared varieties, n_x represents the number of all bands of one of the compared varieties, and n_y represents the number of all bands of the other compared variety.

Separately for each of the indices, these coefficients of similarity were used for calculating a pairwise distance matrix for each marker, where the distance between two selected varieties was computed as $1 - \text{the corresponding similarity coefficient}$. Finally, an averaged distance matrix was created by averaging distance matrices of all markers. Thus, two pairwise distance matrices—one based on the Jaccard index and one on the Sørensen–Dice index—were created, describing the averaged dissimilarity between each pair of red clover varieties. Two phylogenetic trees based on the averaged distance matrices were calculated in MATLAB (version R2015a, <http://www.mathworks.com>) using the unweighted pair group method with arithmetic mean (UPGMA) clustering method then manually edited and visualized in FigTree (version 1.4.2, <http://tree.bio.ed.ac.uk/software/figtree/>).

RESULTS

Sequencing and Gene Annotation

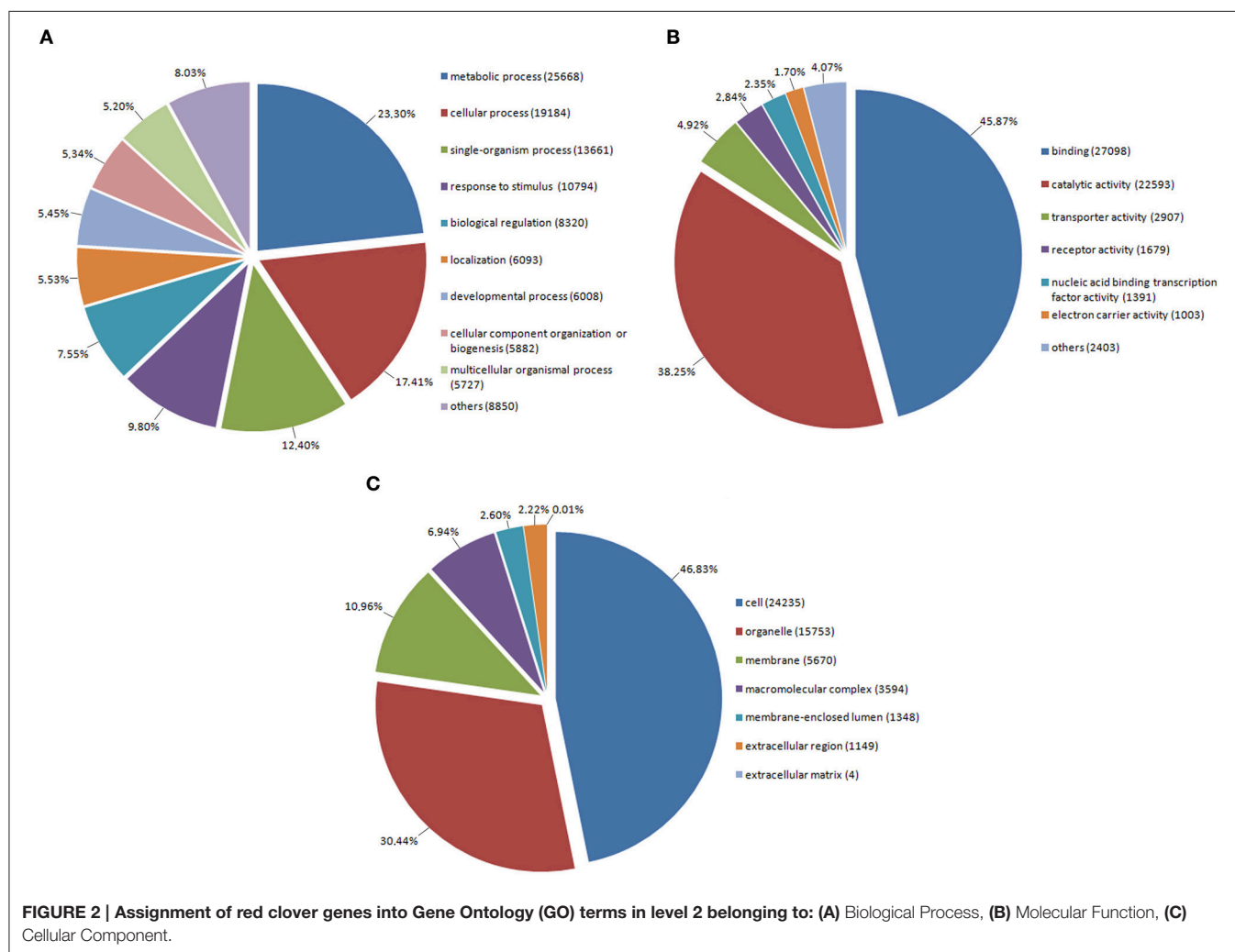
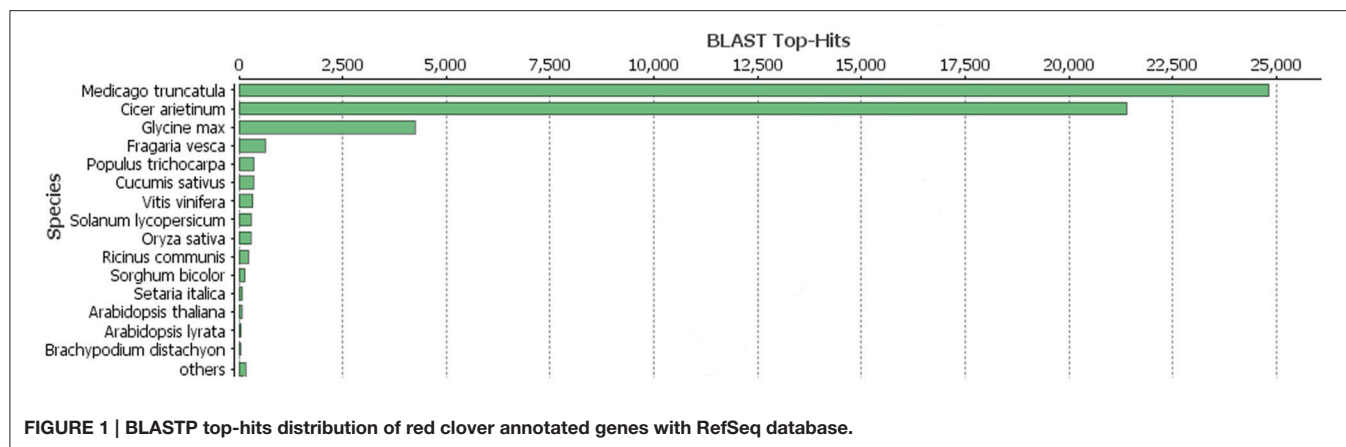
As described in Ištvánek et al. (2014), 243.6 million reads were obtained by sequencing. After filtering out low-quality reads and sequencing adapter relics, genome coverage of $\sim 55.4\times$ was achieved. A total of 64,761 genes were predicted in red clover (Ištvánek et al., 2014) and after manual revision of Blast2GO annotation, 42,996 genes were characterized. These included 1,316 genes related to repetitive elements (Table 2). One of the main annotation steps was based on finding the sequence homology with accessions in the RefSeq database (BLASTP search). The results of this part are summarized in Figure 1 in the form of Blast Top-Hits, showing the degree of relationship to other sequenced plant model species. All predicted genes with their annotations are displayed in Table S5.

Annotated genes were assigned to appropriate biological process, molecular function, and cell component subclasses based on their annotation (Figure 2). Within the sequences associated with biological processes, GO terms associated with primary and secondary metabolism were the most prevalent. In this respect, primary metabolites are known to be essential for plant survival while secondary metabolites play important roles in plant

TABLE 2 | Red clover gene characteristics.

PREDICTED GENES	
Number of genes	64,761
Number of exons	202,783
Number of introns	160,364
Mean number of exons per gene	3.1
Mean number of introns per gene	2.5
Mean gene length (bp)	1,480.3
Mean exon length (bp)	244.6
Mean intron length (bp)	288.5
Single-exon genes	11,559 (17.8%)
ANNOTATED GENES	
Number of genes	42,996
Number of exons	160,421
Number of introns	131,368
Mean number of exons per gene	3.7
Mean number of introns per gene	3.1
Mean gene length (bp)	1,818.1
Mean exon length (bp)	244.9
Mean intron length (bp)	294.6
Single-exon genes	6,854 (15.9%)
GENES WITHOUT ANNOTATION	
Number of genes	21,765
Number of exons	42,362
Number of introns	28,396
Mean number of exons per gene	1.9
Mean number of introns per gene	1.3
Mean gene length (bp)	813.0
Mean exon length (bp)	243.3
Mean intron length (bp)	260.1
Single-exon genes	4,705 (21.6%)
GENES RELATED TO REPETITIVE ELEMENTS	
Number of genes	1,316
Number of exons	2,360
Number of introns	1,382
Mean number of exons per gene	1.8
Mean number of introns per gene	1.1
Mean gene length (bp)	1,359.0
Mean exon length (bp)	655.6
Mean intron length (bp)	174.6
Single-exon genes	301 (22.9%)

protection and have a broad spectrum of utilization. We also found genes associated with the GO term “response to stimulus” to occur very frequently. This category includes mainly genes involved in responses to stress, biotic stress, and endogenous as well as extracellular stimuli. In molecular function, almost one-half of genes are associated with the GO term “binding,” in which the binding functions of nucleotides or DNA form the majority. “Catalytic activity,” as the second most prevalent term, comprises enzymatic activities of kinases, hydrolases, nucleases, etc. In cellular component, the most frequent GO terms were associated with functions within the plant cell, organelles (plastids and mitochondria), and plasma membrane. In short, these relate to the main cellular compartments of plant cells.



Annotated genes were also classified into metabolic and biosynthetic pathways. A total of 7,517 genes were characterized and assigned to specific pathways. Because many genes figure in multiple biosynthetic or metabolic pathways, a total of 17,727 enzymatic nodes were described in all pathways.

Among the largest metabolic pathways (each involving more than 1,000 genes) were purine metabolism and starch and sucrose metabolism. The 20 largest biosynthetic pathways are summarized in **Table 3**. Table S6 presents a complete list of genes assigned to specific metabolic and biosynthetic

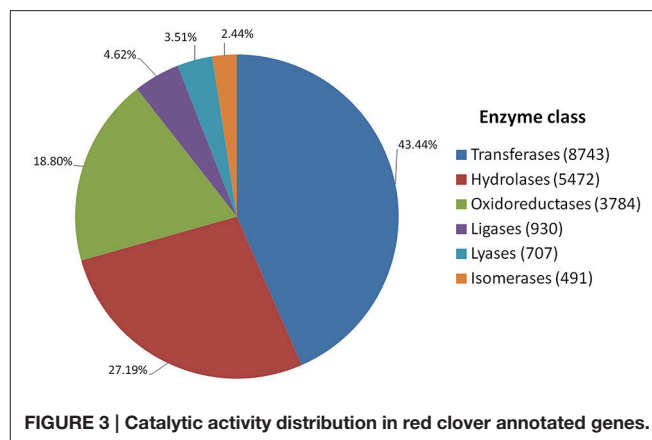
TABLE 3 | Twenty largest biosynthetic and metabolic pathways in red clover based on number of genes (enzymes) involved.

Pathway	Seqs. in pathway
Purine metabolism	1,138
Starch and sucrose metabolism	1,053
Phenylalanine metabolism	622
Pentose and glucuronate interconversions	475
Phenylpropanoid biosynthesis	469
Thiamine metabolism	428
Pyrimidine metabolism	378
Glycerolipid metabolism	364
Cysteine and methionine metabolism	354
Galactose metabolism	328
Amino sugar and nucleotide sugar metabolism	323
Glycerophospholipid metabolism	304
T cell receptor signaling pathway	297
Glycolysis/Gluconeogenesis	284
Tyrosine metabolism	273
Phenylalanine, tyrosine and tryptophan biosynthesis	272
Flavonoid biosynthesis	257
Arginine and proline metabolism	253
Glutathione metabolism	242
Pyruvate metabolism	240

pathways. Each enzyme was also assigned to one of the main enzyme classes (Figure 3). In red clover, almost one-half of enzymes belong to transferases (43.44%), and more than 89% of enzymes consist of transferases, hydrolases, and oxidoreductases.

Comparison of Red Clover Genes with Model Legumes

A TBLASTX search was performed to evaluate the distribution of all red clover genes along recently published chromosomes of red clover, chromosomes of the model legume species *M. truncatula* (8 chromosomes), and chromosomes of *G. max* (20 chromosomes). The results were plotted using a window size of 100 kb through genomic sequences (Figure 4). Repetitive element and gene densities in each species were distributed along all chromosomes of *T. pratense*, *M. truncatula*, and *G. max*. Distribution patterns were similar in both *T. pratense* and *M. truncatula*. Under the specified criteria, 41,607 red clover genes were found to be homologs in comparison with *M. truncatula* and 32,737 genes were homologs with *G. max*. In *G. max*, the genes were concentrated in subtelomeric and telomeric regions. These are regions with low density of repetitive elements, unlike centromeric regions (Torales et al., 2013). This can be seen also in the central lines that show the distribution of homologous sequences to *M. truncatula*, *G. max*, and red clover. On the other hand, the gene densities in *M. truncatula* are the more balanced, with only a slight decrease in centromeric regions. Centromeric regions were also poorer for homologous sequences, for example in chromosome Mt6 and Mt8.

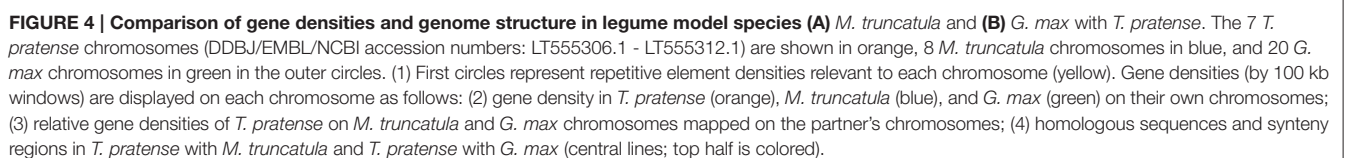
**FIGURE 3 | Catalytic activity distribution in red clover annotated genes.**

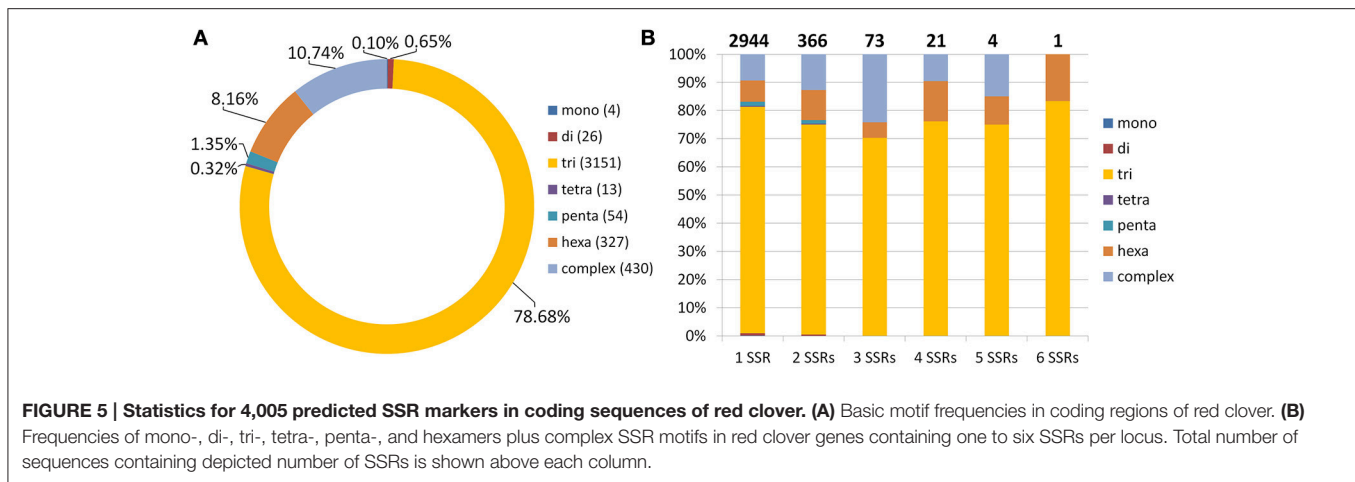
Predicted DNA Markers

Using SSR Locator (da Maia et al., 2008), we identified 6,749 potential microsatellite loci in red clover coding sequences. For those with sufficient flanking sequences, we designed appropriate unique primers to generate PCR product preferentially within 100–350 bp. The resulting 4,005 (59.3%) potential SSR markers were characterized (Figure 5). Because 1,061 (26.5%) of these SSR markers occurred in an identical unique locus, it results that just 3,409 (5.3%) coding sequences possess at least one SSR marker. When the total length of coding sequences (49.6 Mbp) is taken into count, marker density of 1 SSR marker per 12.39 kbp was achieved. Especially noteworthy is that no SSR markers were found in the genes belonging to the isoflavonoid biosynthetic pathway, such as 2-dihydroflavonol reductase, chalcone synthase, and isoflavone synthase. All potential SSR markers are shown in Table S7.

As expected, the most frequently seen basic motif of microsatellite corresponded to trimeric repeat (78.68%), followed by complex (10.74%) and hexameric (8.16%) motifs (Figure 5). These motifs were also present mainly in loci with a single SSR marker. Complex motifs consisted mainly of two–five trimeric motifs, with only 55 (12.8%) exceptions containing also other motifs (mainly hexameric). Other motifs, such as dimeric and pentameric, were seen much less frequently. Only in 7 SSRs with complex motifs did the complex motif not contain a trimeric repeat.

SNPs were identified by aligning reads to predicted coding sequences. The analysis resulted in identification of 343,027 SNP markers, providing marker density of 1 SNP marker per 144.6 bp, meaning on average 5.3 SNP markers per gene. Of these SNPs, 290,905 (84.8%) SNPs were high quality. SNP markers were also divided between transitions and transversions based on the nature of the A allele. The majority (nearly two-thirds) of SNP markers were transitions. In addition, 4,065 (1.19%) of the identified SNP markers were multi-allelic, with more than one A allele. Table 4 presents a complete overview and statistics relating to SNP markers. Table S8 summarizes the complete list of SNP markers, including their positions and additional information.





Validation and Polymorphism of Predicted SSR Markers, Phylogenetic Analysis

Of 96 chosen SSR loci, only the SSR locus SSR-TP_g53834.t1.cds1 was not amplified. Altogether, 95 SSRs were analyzed for 50 red clover varieties. SSR markers with a PCR product are summarized in Table S1. Single monomorphic SSR marker SSR-TP_g20700.t1.cds3 was amplified in all 50 varieties. The lowest number of amplified samples/varieties was 27 (**Figure 6**). Fifteen varieties gave a PCR product for all 95 SSRs, and the lowest number of markers (4) was amplified in a single variety (Radegast).

Allele number ranged from 1 to 17 (Table S1, **Figure 6**). The pPIC of these SSR loci ranged from 0 to 0.986 with a mean of 0.679 and median of 0.693 (Table S1). The highest diversity was determined for SSR loci with trinucleotide motifs and pPIC ranging from 0.180 to 0.986 (**Table 5**). Twenty-two SSRs out of 95 validated were highly polymorphic and with pPIC > 0.9 while 72 SSRs showed pPIC > 0.5 (**Figure 6**).

The similarity between individual varieties of *T. pratense* was assessed using the Sørensen–Dice and Jaccard indices (**Figure 7**, Figure S1). Cluster analysis grouped the 50 red clover varieties into two clusters. Sub-cluster IA consisted of the single variety Radegast (4x), developed from landraces that were well adapted locally, from breeding varieties (Slovensky podtatransky, Chlumecky, and Horal) and a later cross with the variety Weitetra. Sub-cluster IIA comprised a group of varieties whose genomes were enriched with genotypes of European origin: Dolina (4x), Vulkan (4x), and Sigord (4x) were developed by crosses of Czech, Polish, and German varieties; Tabor (2x) was developed by mass crosses of selected resistant plants belonging to 49 varieties; and Atlantis (4x) is of German origin. Cluster IIB1 consisted of two varieties: Slavoj (2x) was developed by the selection of genotypes, and Kvarta (4x; released 1974) was developed by polyploidy of landraces and the variety Chlumecky (2x), which itself was a component of the next sub-cluster IIB2-1. Chlumecky is the earliest red clover cultivar (released 1935) developed by individual plant selection from the landrace Cesky. Sprint (4x) was obtained after the polyploidy of four newly bred genotypes of European origin.

IIB2-2 was a large cluster of diploid and tetraploid Czech, European, and non-European varieties, reflecting that the varieties are often populations developed from genotypes with wide genetic variability, by targeted crosses, polycrosses, and topcrosses suitable for the selection of complex characters, with synteny of the selected genotypes. The sub-clusters of non-Czech origin varieties were as follows: IIB2-2a Grasslands Hamua from New Zealand and Hungarotetra from Hungary; IIB2-2b three varieties of non-European origin (Makimidori, Concorde, Walter); and IIB2-2c, five European varieties (Lossam, Triton, Essex Broad Red, Gibrinij Pozdnespelyj, Parka). In addition, Pavo and Astur were released in Switzerland, Nemaro, and Titus in Germany, and Vesna is of Czech origin but developed by crosses of diploid genotypes from Czech, French, Swiss, and German varieties with subsequent polyploidy by colchicine. Slovak and Swedish red clover genetic material was introgressed into the genome of Tatra, and Blizard was bred using non-Czech genotypes and recurrent phenotypic selection. Start, released in 1974, was used as a component for more recently bred varieties such as Garant, Cyklon, Spur, Trubadur, Dolly, and Tempus. The same cluster distribution was observed using the Sørensen–Dice and Jaccard indices, with only a few exceptions in cluster IIB2-2 (**Figure 7**, Figure S1), such as the sub-clusters Grasslands Hamua and Hungarotetra, Cyklon, Suez, Beskyd, Spur, and Trubadur.

Validation of SNP Markers and Their Polymorphism

We examined intra-variety genetic heterogeneity using genome-wide SNP genotyping. Five possible genotypes for two alleles per SNP (reference *R*, alternative *A*) were differentiated (RRRR, RRRA, RRAA, RAAA, AAAA) for tetraploid plants. Our analysis revealed 8,607 polymorphic SNP markers with PIC ranging from 0.024 to 0.375 with a mean of 0.338 and median of 0.355 (Table S9).

Single-marker polymorphism was successfully validated and confirmed in five SNP loci and homozygosity/heterozygosity was determined in 14 particular plants, 7 of which were of variety Tatra, 4 of variety Tempus, and 1 plant of each Start, Amos, and Fresco. The majority of those plants analyzed were heterozygous

TABLE 4 | Statistical overview of SNP markers predicted in red clover.

SNP	Transitions		SNP	Transversions	
	Number	Prevalence %		Number	Prevalence %
A <-> G	109,284	50.15	A <-> C	32,387	26.75
C <-> T	108,610	49.85	C <-> G	21,276	17.57
Total	217,894	63.52	G <-> T	29,231	24.14
			T <-> A	38,174	31.53
Multi-allelic SNP	4,065	1.19	Total	121,068	35.29

in the tested loci. Homozygosity for the *R* allele was detected in one plant (Tempus) in TP_g30014_516, four plants (Tatra, Amos, Fresco, Tempus) in TP_g30658_273, two plants (Tatra, Tempus) in TP_g33120_639, two plants (Amos, Tempus) in TP_g51879_538, and eight plants (Tatra, Start, Amos, Fresco, and all Tempus) in TP_g56325_406. Two homozygotes of Tatra were detected for the *A* allele in TP_g33120_639 (**Figure 8**). All amplified fragments corresponded to those predicted.

DISCUSSION

T. pratense Genes

The number of annotated genes in this study is higher than the number of genes identified recently by RNA sequencing (34,534 genes; Yates et al., 2014), but it is very close to the number from other WGS (40,868 genes; De Vega et al., 2015). On the basis of improved annotation of red clover genes, genes were classified into biosynthetic and metabolic pathways and key enzymes were identified. We have found 1,138 genes involved in purine metabolism, which is the fundamental pathway for plant growth and development (Zrenner et al., 2006). This pathway is associated with DNA synthesis, energy sources, and synthesis of many primary and secondary metabolic products (Stasolla et al., 2003). More than 1,000 genes are also involved in starch and sucrose metabolism, which is one of the most important pathways regarding energy sources in plants. Moreover, with 257 genes involved, biosynthesis of flavonoids is among the largest metabolic and biosynthetic pathways. These genes are of particular interest to red clover breeders inasmuch as flavonoid biosynthesis is associated with isoflavonoid content in the plant. Because they are known plant estrogen analogs, high levels of isoflavonoids are undesirable in forage varieties (Adams, 1995) but these are required in varieties used in pharmaceuticals (Park and Weaver, 2012).

Based on the distribution of homologous sequences among red clover, *M. truncatula*, and *G. max*, very similar patterns in gene distribution are present in red clover and *M. truncatula*. In contrast to *G. max*, gene density is rather uniform along entire chromosomes. Occasional spikes in density of red clover genes show clusters of numerous gene families, such as genes of resistance. As previously described (Kulikova et al., 2004; Isobe et al., 2012; Torales et al., 2013), *M. truncatula* had fewer homologous genes located on its chromosome 6 due to the abundance of heterochromatic sites and retroviral elements

scattered throughout its arms. This was in contrast to red clover, where no such decrease was observed. The density of repetitive content along the chromosomes also was inspected and this provided very similar results. Also visible, however, is increased content of repetitive elements in red clover (Ištvánek et al., 2014) compared to *M. truncatula* (Young et al., 2011). Orthologous loci were connected and compared among species visualizing syntenic loci and rearrangements of genome structure. During speciation, red clover clearly underwent complex genome restructuring, possibly associated with reduction of the basic chromosome number from eight to seven. Results supporting this hypothesis were also found in a comparison among red clover, white clover (*T. repens*), and *M. truncatula* based on comparing DNA markers and their location in the genomes (Isobe et al., 2012). Our results are supported, too, by a recently published paper regarding WGS of red clover and construction of its physical map (De Vega et al., 2015). A slight discrepancies in locations of homologous sequences are likely the result of different methodology compared to previously published papers (Isobe et al., 2012; De Vega et al., 2015).

DNA Markers

DNA markers have a broad spectrum of uses in both research and practical breeding. They are used in QTL mapping (Zhao et al., 2013), evolution relationship studies (Ghamkhar et al., 2012; Isobe et al., 2012), variability assessment and genotyping of breeding material (Younas et al., 2012; Cidade et al., 2013), marker-assisted selection, and even gene pyramiding (Qi et al., 2015). Based on NGS technology, we are capable of discovering thousands of SSR and many millions of SNP markers (Zalapa et al., 2012).

Searching for SSR loci within 64,761 predicted coding sequences resulted in the identification of 6,749 SSR loci. This is more than twice the number of SSR loci identified from red clover transcriptome sequencing (Yates et al., 2014). On the other hand, due to the lower number of genes and shorter length of coding sequence described by RNA sequencing, the average SSR marker frequency is very similar (1 SSR marker per 13.42 kbp; Yates et al., 2014). When multiple repeat occurrences are taken into account, 10.4% of genes on average contained an SSR locus. This frequency is comparable to that reported in *Prosopis alba* (11%; Torales et al., 2013) but lower than those in *Nothofagus nervosa* (15%; Torales et al., 2012) and oak (19%; Ueno et al., 2010 and 24%; Durand et al., 2010). In coding sequences, clear domination of trimeric motifs (78.68%) is observed when compared to the SSRs from the genome as a whole (26.9%; Ištvánek et al., 2014). Much lower frequencies were found for other motifs. Similar results have been obtained also in other species, such as *N. nervosa* (Torales et al., 2012) and oak (Ueno et al., 2010). This phenomenon is very likely connected to the need to preserve open reading frame within the coding sequences and negative selection pressure against those SSR loci breaking it. Even in the majority of SSR loci with non-trimeric basic motifs, therefore, the combination of motif length and its repeat number is divisible by three—e.g., (A)₁₂, (GA)₆, (ATTGG)₃—and this, then, does not violate reading frame by frame-shift mutations (Metzgar et al., 2000).

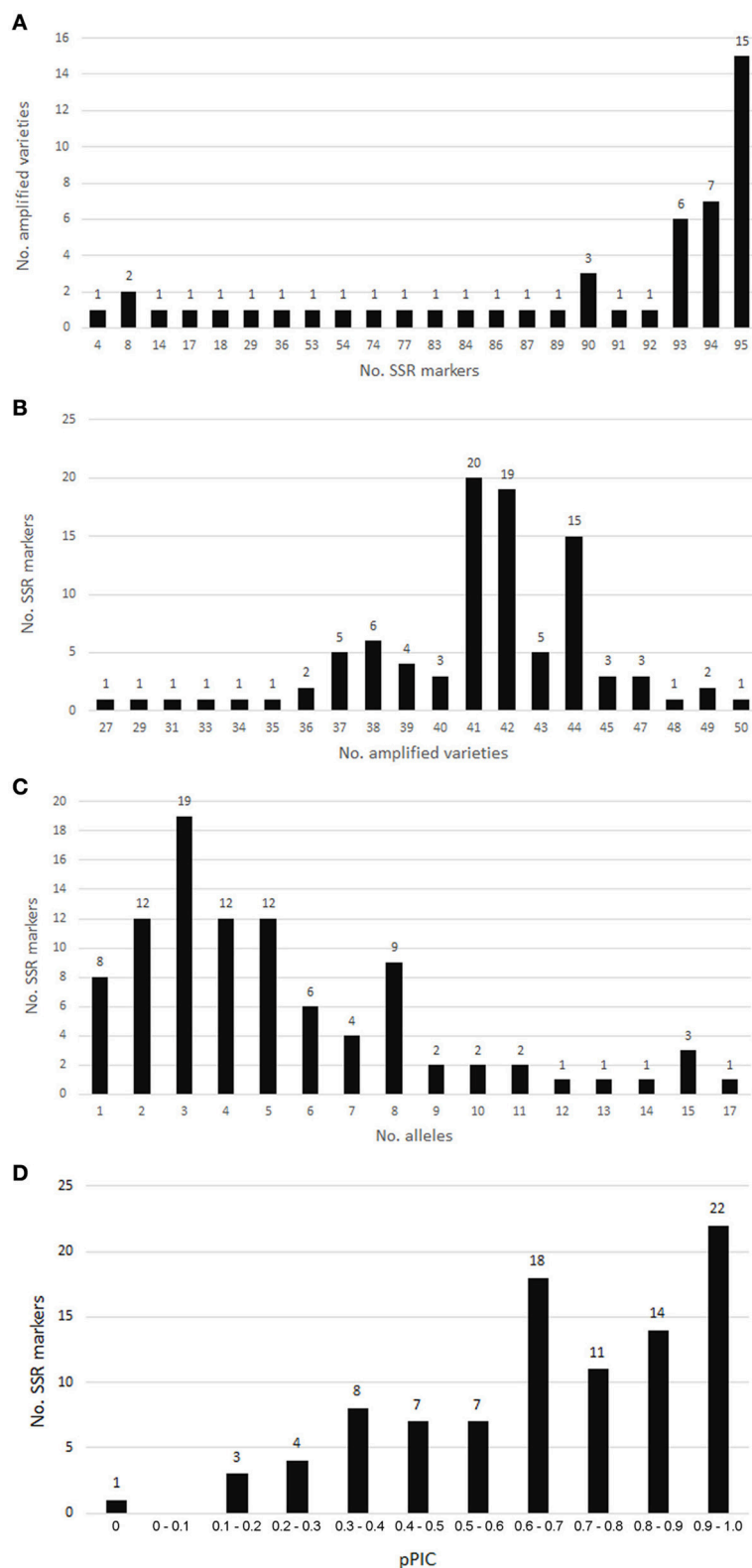


FIGURE 6 | SSR marker validation in red clover varieties. (A) Numbers of SSRs amplified in analyzed varieties. **(B)** Numbers of varieties amplified for individual SSRs. **(C)** Allele number distribution for SSR markers validated in red clover varieties. **(D)** pPIC distribution for SSR markers validated in red clover varieties.

SNPs were also identified in coding sequences of red clover. Their greater frequency throughout the genome (343,027 SNP markers; 1 SNP per 144.6 bp) makes them useful in such high-throughput methods as SNP arrays (Viquez-Zamora et al., 2013; Yu et al., 2014). Nevertheless, when compared to other plant species, it is clear that SNP frequency is influenced by many factors, such as the number of individual plants analyzed in the study, natural variability in the population of the studied species, etc. In *P. alba*, for example, 1 SNP marker was found for every 2,512 bp (Torales et al., 2013), in *Capsicum annuum* 1 for every 2,253 bp (Ashrafi et al., 2012), in oak 1 for every 471 bp (Ueno et al., 2010), and in *Eucalyptus grandis* 1 for every 192 bp (Novaes et al., 2008). In these studies, the SNP number found correlated mainly with the number of individuals analyzed (e.g., 21 individual plants in oak, more than 200 in *E. grandis*). Although, just 16 individual plants of the same variety were analyzed in red clover, even higher SNP frequency was obtained, likely due to the outcrossing nature of clovers. Within the identified SNPs, transitions (63.52%) showed significant dominance over transversions (35.29%). These results are consistent with those in *P. alba* (Torales et al., 2013) and *Cucurbita pepo* (Blanca et al., 2011). A large number of SNPs are now available in red clover for genome-wide association studies and SNP microarray construction, where tens of thousands of markers are required.

Validation of DNA Markers and Polymorphism

Outcrossing species populations are exceptionally variable and with a high level of heterozygosity. The majority of genetic analyses of such species are necessarily carried out in pooled samples in order to collect most of the population variability and also minimize costs. Results obtained from such pooled samples are, however, unsuitable for estimating the copy number of individual alleles, which precludes assessment of exact allele frequencies required to calculate polymorphic information content (PIC; Botstein et al., 1980). Recent advances in NGS technologies enable determination of allele frequencies from pooled samples (Mullen et al., 2012; Lynch et al., 2014), but these are very expensive and thus inaccessible especially for breeders working with non-model crops.

PIC is commonly used in plant genetics to assess polymorphism level for a marker locus. For leguminous plants, PIC was recently used, for example, for evaluating 48 SSR markers of *Vigna radiata* (Shrivastava et al., 2014), 45 SSR markers of *Trifolium alexandrinum* (Verma et al., 2015), and 36 SSR markers of *Vicia* spp. (Raveendar et al., 2015). Estimated PIC is usually directly connected with suitability for subsequent utilization, such as in variety identification or selection of suitable material for breeding purposes. In order to calculate PIC, a precise determination of allelic frequencies in the studied population is required.

To overcome the disadvantages of pooled samples, we proposed a modified PIC value termed pooled polymorphic information content (pPIC) which does not rely on determining allelic frequencies in the selected population. pPIC ranges from

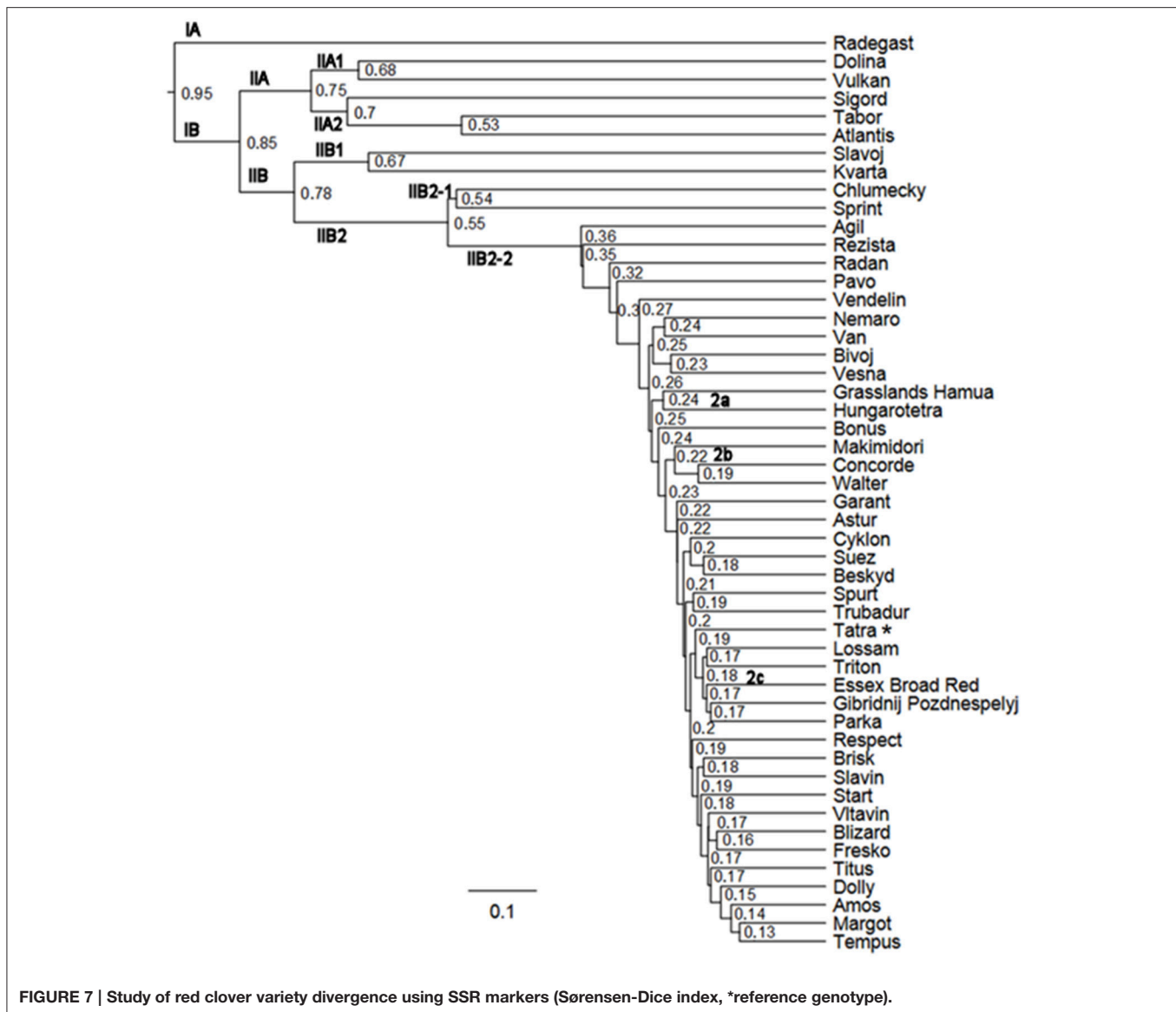
TABLE 5 | pPIC in validated red clover SSRs with different motifs.

Repeat type	Number of SSRs	pPIC		
		Range	Mean	SD
Monomer	2	0.567–0.693	0.6297	0.0632
Dimer	2	0.875–0.895	0.8853	0.0102
Trimer	67	0.180–0.986	0.6929	0.2227
Hexamer	2	0.333–0.365	0.3488	0.0160
Complex	22	0–0.975	0.6514	0.2734

0 to 1, similarly as does PIC, and it estimates the probability that two randomly collected pooled samples from the given species will differ in the given marker. Unlike PIC, pPIC works well for assessing the polymorphism level of a marker locus for pooled samples. The presented pPIC of the 95 SSR markers analyzed should, however, be taken into account only to evaluate pooled samples similar in size to that of our study. A significant decrease or increase in pooled sample size could shift pPIC and thus degrade the estimation of SSR marker discrimination power. Our results based on a pooled sample size of 16 plants should nevertheless be optimal for most potential subsequent utilizations. This is particularly important for screening gene bank accessions and large-scale analysis of cultivar identity and seed purity. For red clover, moreover, the optimal bulk size for genetic variation assessment among cultivars has been determined as 20 (Kongkiatngam et al., 1996).

This study generated a collection of 22 highly polymorphic SSRs with pPIC >0.9 and thus primer pairs for application to diversity studies in *T. pratense*. Seventy-two SSRs out of 95 validated showed pPIC >0.5. The single SSR marker in coding sequences SSR-TP_g20700.t1.cds3 was amplified in all 50 varieties but was also monomorphic (pPIC = 0.0). All other SSRs revealed some polymorphism in the analyzed variety populations. For the validated SSRs, the actual length range of amplified fragments corresponded with expectations with the single exception of SSR-TP_g32548.t1.cds1, whose amplified fragment was shorter (100–200 bp) than expected (237 bp).

The breeding methods in red clover include procedures suitable for outcrossing crops. Useful variation in a breeding population can be generated through hybridization and genome introgression, or by chromosome doubling (polyploidy) by colchicine. Subsequent phenotypic selection of superior individual plants or mass selection must be conducted on the progeny combining the best traits, and successive population breeding is performed. Molecular characterization of the analyzed varieties using SSRs reflects their genetic relationships, and the grouping is shown in **Figure 7** and Figure S1. Tracing the breeding history revealed frequent sharing and exchange of cultivars and newly bred materials among European breeding stations. It was shown that varieties from sub-clusters IA, IIA, IIB1, and IIB2-1 had higher relatedness than varieties from sub-cluster IIB2-2. The possible reasons could be (i) introgressions from landraces and (ii) that the varieties were mostly released from 1970 to 1990, Chlumecky as early as 1935 (with the

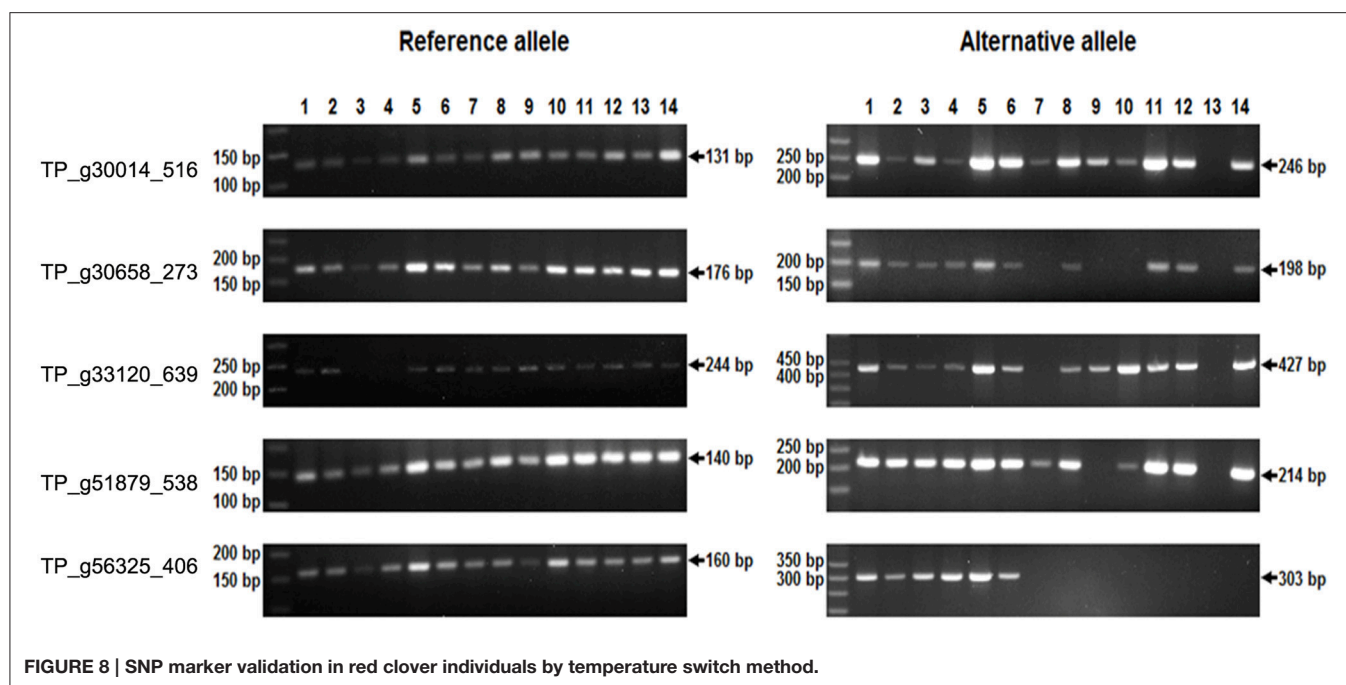


exceptions of Atlantis and Slavoj). Narrowing of the genetic base in the more recent varieties in sub-cluster IIB2-2 was also apparent.

The SSR profiling (i) differentiated varieties with possible introgressions from landraces and (ii) indicated the existence of diversity at the molecular level among different red clover varieties. The finding of inter-variety heterogeneity has important consequences for breeders who use these varieties. Cluster analysis by means of DNA profiling using the validated SSR set is suitable for such study.

Further progress in red clover breeding can be made by crosses with more distant genotypes as sources of new genetic variability, with new introgressions of important loci for resistance and quality. The identification of SSR or SNP markers in known-function genes linked to specific traits can facilitate marker-assisted selection. One important task is to develop a platform

for red clover genotyping, employing genome-wide distributed SNP markers. The Tatra-derived reference sequence was initially used for the detection of the predicted 343 thousand SNPs. We used a preliminary set of 8,623 genome-wide distributed SNPs for polymorphism evaluation in individual plants. Arrayit methods provide universal microarray-based platforms for SNP genotyping (Schena et al., 1996). Sixteen of the validated SNPs were monomorphic and 8,607 were polymorphic with a mean PIC of 0.338. SNP validation confirmed the high quality of SNPs chosen for microarray. More sequenced red clover varieties/genotypes and a large set of informative SNPs are greatly needed for genotyping and association study. NGS methods such as genotyping-by-sequencing and the resequencing of targeted DNA regions from contrasting genotypes appear to be the most essential for SNP discovery and genotyping applications in red clover breeding. Temperature switch PCR can be successfully



used in diagnostic applications through single-marker SNP genotyping for targeted coding sequences and for heterozygosity or homozygosity confirmation in validated loci. Large SNP sets are already available in grain legumes such as soybean (Song et al., 2013; Lee et al., 2015) and pea (Sindhu et al., 2014; Tayeh et al., 2015), or in peanut (Pandey et al., 2017). High-density SNP microarrays can significantly advance breeding applications.

AUTHOR CONTRIBUTIONS

Jl, JN, and JR designed the study. Jl processed sequencing data, characterized protein-coding genes, and collaboratively with JD performed detailed inspection of gene annotation manually. Jl performed gene classification into metabolic and biosynthetic pathways, comparison with other legumes, and generated genome-wide SSR and SNP markers. LP and JD prepared biological material, performed DNA isolation and marker validation. JD and PD performed pPIC and PIC calculation, polymorphism evaluation and phylogenetic analysis.

JR supervised all aspects of the presented analyses. All of the authors contributed to the writing of the manuscript.

ACKNOWLEDGMENTS

The authors thank the Ministry of Agriculture of the Czech Republic (grant no. QI111A019) for financial support. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme “Projects of Large Research, Development, and Innovations Infrastructures.” Seeds were procured from the GeneBank of Crop Research Institute Ltd., Prague-Ruzyně, Czech Republic.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00367/full#supplementary-material>

REFERENCES

- Adams, N. R. (1995). Detection of the effects of phytoestrogens on sheep and cattle. *J. Anim. Sci.* 73, 1509–1515. doi: 10.2527/1995.7351509x
- Ashrafi, H., Hill, T., Stoffel, K., Kozik, A., Yao, J., Chin-Wo, S. R., et al. (2012). *De novo* assembly of the pepper transcriptome (*Capsicum annuum*): a benchmark for *in silico* discovery of SNPs, SSRs and candidate genes. *BMC Genomics* 13:571. doi: 10.1186/1471-2164-13-571
- Blanca, J., Cañizares, J., Roig, C., Ziarsolo, P., Nuez, F., and Picó, B. (2011). Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae). *BMC Genomics* 12:104. doi: 10.1186/1471-2164-12-104
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.
- Cidade, F. W., Vigna, B. B., de Souza, F. H., Valls, J. F., Dall'Agnol, M., Zucchi, M. I., et al. (2013). Genetic variation in polyploid forage grass: assessing the molecular genetic variability in the *Paspalum* genus. *BMC Genet.* 14:50. doi: 10.1186/1471-2156-14-50
- da Maia, L. C., Palmieri, D. A., De Souza, V. Q., Kopp, M. M., de Carvalho, F. I. F., and Costa de Oliveira, A. (2008). SSR Locator: tool for simple sequence repeat discovery integrated with primer design and PCR simulation. *Int. J. Plant Genomics* 2008:412696. doi: 10.1155/2008/412696

- De Vega, J. J., Ayling, S., Hegarty, M., Kudrna, D., Goicoechea, J. L., Ergon, A., et al. (2015). Red clover (*Trifolium pratense* L.) draft genome provides a platform for trait improvement. *Sci. Rep.* 5:17394. doi: 10.1038/srep17394
- Dellaporta, S. L., Wood, J., and Hicks, J. B. (1983). A plant DNA miniprep: version II. *Plant Mol. Biol. Rep.* 1, 19–21. doi: 10.1007/BF02712670
- Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology* 26, 297. doi: 10.2307/1932409
- Durand, J., Bodénès, C., Chancerel, E., Frigerio, J.-M., Vendramin, G., Sebastiani, F., et al. (2010). A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* 11:570. doi: 10.1186/1471-2164-11-570
- Forster, J. W., Jones, E. S., Kölliker, R., Drayton, M. C., Dupal, M. P., Guthridge, K. M., et al. (2001). "Application of DNA profiling to an outbreeding forage species," in *Plant Genotyping: The DNA Fingerprinting of Plants*, ed R. J. Henry (Wallingford: CABI), 299–320. Available online at: <http://www.cabi.org/cabebooks/ebook/20083015002> (Accessed February 9, 2016).
- Ghamkhar, K., Isobe, S., Nichols, P. G. H., Faithfull, T., Ryan, M. H., Snowball, R., et al. (2012). The first genetic maps for subterranean clover (*Trifolium subterraneum* L.) and comparative genomics with *T. pratense* L. and *Medicago truncatula* Gaertn. to identify new molecular markers for breeding. *Mol. Breed.* 30, 213–226. doi: 10.1007/s11032-011-9612-8
- Graham, P. H., and Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant Physiol.* 131, 872–877. doi: 10.1104/pp.017004
- Herrmann, D., Boller, B., Studer, B., Widmer, F., and Kölliker, R. (2008). Improving persistence in red clover: insights from QTL analysis and comparative phenotypic evaluation. *Crop Sci.* 48:269. doi: 10.2135/cropsci2007.03.0143
- Isobe, S., Kölliker, R., Hisano, H., Sasamoto, S., Wada, T., Klimenko, I., et al. (2009). Construction of a consensus linkage map for red clover (*Trifolium pratense* L.). *BMC Plant Biol.* 9:57. doi: 10.1186/1471-2229-9-57
- Isobe, S. N., Hisano, H., Sato, S., Hirakawa, H., Okumura, K., Shirasawa, K., et al. (2012). Comparative genetic mapping and discovery of linkage disequilibrium across linkage groups in white clover (*Trifolium repens* L.). *G3* 2, 607–617. doi: 10.1534/g3.112.002600
- Ištvánek, J., Jaroš, M., Křenek, A., and Řepková, J. (2014). Genome assembly and annotation for red clover (*Trifolium pratense*; Fabaceae). *Am. J. Bot.* 101, 327–337. doi: 10.3732/ajb.1300340
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bull. Soc. Vaudoise Sci. Nat.* 37, 547–579. doi: 10.5169/seals-266450
- Jakešová, H., Řepková, J., Nedělník, J., Hampel, D., Dluhošová, J., Soldánová, M., et al. (2015). Selecting plants with increased total polyphenol oxidases in the genus *Trifolium*. *Czech J. Genet. Plant Breed.* 51, 155–161. doi: 10.17221/107/2015-CJGPB
- Jones, B. A., Hatfield, R. D., and Muck, R. E. (1995). Screening legume forages for soluble phenols, polyphenol oxidase and extract browning. *J. Sci. Food Agric.* 67, 109–112. doi: 10.1002/jsfa.2740670117
- Kataoka, R., Hara, M., Kato, S., Isobe, S., Sato, S., Tabata, S., et al. (2012). Integration of linkage and chromosome maps of red clover (*Trifolium pratense* L.). *Cytogenet. Genome Res.* 137, 60–69. doi: 10.1159/000339509
- Klimenko, I., Razgulyeva, N., Gau, M., Okumura, K., Nakaya, A., Tabata, S., et al. (2010). Mapping candidate QTLs related to plant persistency in red clover. *Theor. Appl. Genet.* 120, 1253–1263. doi: 10.1007/s00122-009-1253-5
- Kongkiatngam, P., Waterway, M. J., Coulman, B. E., and Fortin, M. G. (1996). Genetic variation among cultivars of red clover (*Trifolium pratense* L.) detected by RAPD markers amplified from bulk genomic DNA. *Euphytica* 89, 355–361.
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109
- Kulikova, O., Geurts, R., Lamine, M., Kim, D.-J., Cook, D. R., Leunissen, J., et al. (2004). Satellite repeats in the functional centromere and pericentromeric heterochromatin of *Medicago truncatula*. *Chromosoma* 113, 276–283. doi: 10.1007/s00412-004-0315-3
- Lee, Y. G., Jeong, N., Kim, J. H., Lee, K., Kim, K. H., Pirani, A., et al. (2015). Development, validation and genetic analysis of a large soybean SNP genotyping array. *Plant J.* 81, 625–636. doi: 10.1111/tpj.12755
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Lynch, M., Bost, D., Wilson, S., Maruki, T., and Harrison, S. (2014). Population-genetic inference from pooled-sequencing data. *Genome Biol. Evol.* 6, 1210–1218. doi: 10.1093/gbe/evu085
- Metzgar, D., Bytof, J., and Wills, C. (2000). Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.* 10, 72–80. doi: 10.1101/gr.10.1.72
- Mullen, M. P., Creevey, C. J., Berry, D. P., McCabe, M. S., Magee, D. A., Howard, D. J., et al. (2012). Polymorphism discovery and allele frequency estimation using high-throughput DNA sequencing of target-enriched pooled DNA samples. *BMC Genomics* 13:16. doi: 10.1186/1471-2164-13-16
- Novaes, E., Drost, D. R., Farmerie, W. G., Pappas, G. J. Jr., Grattapaglia, D., Sederoff, R. R., et al. (2008). High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genomics* 9:312. doi: 10.1186/1471-2164-9-312
- Pandey, M. K., Agarwal, G., Kale, S. M., Clevenger, J., Nayak, S. N., Sriswathi, M., et al. (2017). Development and evaluation of a high density genotyping 'Axiom_Arachis' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Sci. Rep.* 7:40577. doi: 10.1038/srep40577
- Park, C. Y., and Weaver, C. M. (2012). Vitamin D interactions with soy isoflavones on bone after menopause: a review. *Nutrients* 4, 1610–1621. doi: 10.3390/nu411610
- Qi, L. L., Ma, G. J., Long, Y. M., Hulke, B. S., Gong, L., and Markell, S. G. (2015). Relocation of a rust resistance gene R2 and its marker-assisted gene pyramiding in confection sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 128, 477–488. doi: 10.1007/s00122-014-2446-0
- Raveandar, S., Lee, G.-A., Jeon, Y.-A., Lee, Y. J., Lee, J.-R., Cho, G.-T., et al. (2015). Cross-amplification of *Vicia sativa* subsp. *sativa* microsatellites across 22 other *Vicia* species. *Molecules* 20, 1543–1550. doi: 10.3390/molecules20011543
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rogers, S. O., and Bendich, A. J. (1989). "Extraction of DNA from plant tissues," in *Plant Molecular Biology Manual*, eds S. B. Gelvin, R. A. Schilperoort, and D. P. S. Verma (Dordrecht: Springer), 73–83. Available online at: http://www.springerlink.com/index/10.1007/978-94-009-0951-9_6 (Accessed February 9, 2016).
- RStudio Team (2015). *RStudio: Integrated Development for R*. Boston, MA: RStudio, Inc. Available online at: <http://www.rstudio.com/> (Accessed February 6, 2017).
- Sato, S., Isobe, S., Asamizu, E., Ohmido, N., Kataoka, R., Nakamura, Y., et al. (2005). Comprehensive structural analysis of the genome of red clover (*Trifolium pratense* L.). *DNA Res.* 12, 301–364. doi: 10.1093/dnares/dsi018
- Sato, S., Nakamura, Y., Kaneko, T., Asamizu, E., Kato, T., Nakao, M., et al. (2008). Genome structure of the legume, *Lotus japonicus*. *DNA Res.* 15, 227–239. doi: 10.1093/dnares/dsn008
- Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O., and Davis, R. W. (1996). Parallel human genome analysis: micromicroarray-based expression monitoring of 1,000 genes. *Proc. Natl. Acad. Sci. U.S.A.* 93, 10614–10619.
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Shrivastava, D., Verma, P., and Bhatia, S. (2014). Expanding the repertoire of microsatellite markers for polymorphism studies in Indian accessions of mung bean (*Vigna radiata* L. Wilczek). *Mol. Biol. Rep.* 41, 5669–5680. doi: 10.1007/s11033-014-3436-7
- Sindhu, A., Ramsay, L., Sanderson, L. A., Stonehouse, R., Li, R., Condie, J., et al. (2014). Gene-based SNP discovery and genetic mapping in pea. *Theor. Appl. Genet.* 127, 2225–2241. doi: 10.1007/s00122-014-2375-y
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS ONE* 8:e54985. doi: 10.1371/journal.pone.0054985

- Sørensen, T. (1948). *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. I kommission hos E. Munksgaard. Available online at: <https://books.google.co.in/books?id=rpS8GAAACAAJ>
- Sprent, J. I. (2009). *Legume Nodulation*. Oxford, UK: Wiley-Blackwell (Accessed March 3, 2016).
- Stasolla, C., Katahira, R., Thorpe, T. A., and Ashihara, H. (2003). Purine and pyrimidine nucleotide metabolism in higher plants. *J. Plant Physiol.* 160, 1271–1295. doi: 10.1078/0176-1617-01169
- Tabone, T., Mather, D. E., and Hayden, M. J. (2009). Temperature switch PCR (TSP): robust assay design for reliable amplification and genotyping of SNPs. *BMC Genomics* 10:580. doi: 10.1186/1471-2164-10-580
- Tayeh, N., Aluome, C., Falque, M., Jacquin, F., Klein, A., Chauveau, A., et al. (2015). Development of two major resources for pea genomics: the GenoPea 13.2K SNP array and a high-density, high-resolution consensus genetic map. *Plant J.* 84, 1257–1273. doi: 10.1111/tpj.13070
- Teuscher, E., and Lindequist, U. (2010). *Biogene Gifte: Biologie, Chemie, Pharmakologie, Toxikologie, 3 neu bearb.* Stuttgart: Wiss. Verl.-Ges.
- Torales, S. L., Rivarola, M., Pomponio, M. F., Fernández, P., Acuña, C. V., Marchelli, P., et al. (2012). Transcriptome survey of Patagonian southern beech *Nothofagus nervosa* (=N. alpina): assembly, annotation and molecular marker discovery. *BMC Genomics* 13:291. doi: 10.1186/1471-2164-13-291
- Torales, S. L., Rivarola, M., Pomponio, M. F., Gonzalez, S., Acuña, C. V., Fernández, P., et al. (2013). *De novo* assembly and characterization of leaf transcriptome for the development of functional molecular markers of the extremophile multipurpose tree species *Prosopis alba*. *BMC Genomics* 14:705. doi: 10.1186/1471-2164-14-705
- Ueno, S., Le Provost, G., Léger, V., Klopp, C., Noirot, C., Frigerio, J.-M., et al. (2010). Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* 11:650. doi: 10.1186/1471-2164-11-650
- Varshney, R. K., Ribaut, J.-M., Buckler, E. S., Tuberosa, R., Rafalski, J. A., and Langridge, P. (2012). Can genomics boost productivity of orphan crops? *Nat. Biotechnol.* 30, 1172–1176. doi: 10.1038/nbt.2440
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491
- Verma, P., Chandra, A., Roy, A. K., Malaviya, D. R., Kaushal, P., Pandey, D., et al. (2015). Development, characterization and cross-species transferability of genomic SSR markers in berseem (*Trifolium alexandrinum* L.), an important multi-cut annual forage legume. *Mol. Breed.* 35, 1–14. doi: 10.1007/s11032-015-0223-7
- Viquez-Zamora, M., Vosman, B., van de Geest, H., Bovy, A., Visser, R. G., Finkers, R., et al. (2013). Tomato breeding in the genomics era: insights from a SNP array. *BMC Genomics* 14:354. doi: 10.1186/1471-2164-14-354
- Vižintin, L., Javornik, B., and Bohanec, B. (2006). Genetic characterization of selected *Trifolium* species as revealed by nuclear DNA content and ITS rDNA region analysis. *Plant Sci.* 170, 859–866. doi: 10.1016/j.plantsci.2005.12.007
- Wink, M. (2013). Evolution of secondary metabolites in legumes (Fabaceae). *South Afr. J. Bot.* 89, 164–175. doi: 10.1016/j.sajb.2013.06.006
- Yates, S. A., Swain, M. T., Hegarty, M. J., Chernukin, I., Lowe, M., Allison, G. G., et al. (2014). *De novo* assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genomics* 15:453. doi: 10.1186/1471-2164-15-453
- Younas, M., Xiao, Y., Cai, D., Yang, W., Ye, W., Wu, J., et al. (2012). Molecular characterization of oilseed rape accessions collected from multi continents for exploitation of potential heterotic group through SSR markers. *Mol. Biol. Rep.* 39, 5105–5113. doi: 10.1007/s11033-011-1306-0
- Young, N. D., Debellé, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625
- Yu, H., Xie, W., Li, J., Zhou, F., and Zhang, Q. (2014). A whole-genome SNP array (RICE6K) for genomic breeding in rice. *Plant Biotechnol. J.* 12, 28–37. doi: 10.1111/pbi.12113
- Zalapa, J. E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., et al. (2012). Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. *Am. J. Bot.* 99, 193–208. doi: 10.3732/ajb.1100394
- Zhao, P., Zhang, G., Wu, X., Li, N., Shi, D., Zhang, D., et al. (2013). Fine mapping of RppP25, a southern rust resistance gene in maize. *J. Integr. Plant Biol.* 55, 462–472. doi: 10.1111/jipb.12027
- Zrenner, R., Stitt, M., Sonnewald, U., and Boldt, R. (2006). Pyrimidine and purine biosynthesis and degradation in plants. *Annu. Rev. Plant Biol.* 57, 805–836. doi: 10.1146/annurev.arplant.57.032905.105421

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Ištváněk, Dluhošová, Dluhoš, Pátková, Nedělník and Řepková. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Genotyping-by-Sequencing and Its Exploitation for Forage and Cool-Season Grain Legume Breeding

Paolo Annicchiarico^{1*}, Nelson Nazzicari¹, Yanling Wei^{1,2}, Luciano Pecetti¹ and Edward C. Brummer²

¹ Centro di Ricerca per le Produzioni Foraggere e Lattiero-Casearie (FLC), CREA, Lodi, Italy, ² Plant Breeding Center, Department of Plant Sciences, University of California, Davis, Davis, CA, USA

OPEN ACCESS

Edited by:

Oswaldo Valdes-Lopez,
National Autonomous University
of Mexico, Mexico

Reviewed by:

Hamid Khazaei,
University of Saskatchewan, Canada
Bunyamin Tar'an,
University of Saskatchewan, Canada

*Correspondence:

Paolo Annicchiarico
paolo.annicchiarico@crea.gov.it

Specialty section:

This article was submitted to
Crop Science and Horticulture,
a section of the journal
Frontiers in Plant Science

Received: 29 December 2016

Accepted: 13 April 2017

Published: 09 May 2017

Citation:

Annicchiarico P, Nazzicari N, Wei Y,
Pecetti L and Brummer EC (2017)
Genotyping-by-Sequencing and Its
Exploitation for Forage
and Cool-Season Grain Legume
Breeding. *Front. Plant Sci.* 8:679.
doi: 10.3389/fpls.2017.00679

Genotyping-by-Sequencing (GBS) may drastically reduce genotyping costs compared with single nucleotide polymorphism (SNP) array platforms. However, it may require optimization for specific crops to maximize the number of available markers. Exploiting GBS-generated markers may require optimization, too (e.g., to cope with missing data). This study aimed (i) to compare elements of GBS protocols on legume species that differ for genome size, ploidy, and breeding system, and (ii) to show successful applications and challenges of GBS data on legume species. Preliminary work on alfalfa and *Medicago truncatula* suggested the greater interest of *ApeKI* over *PstI*:*MspI* DNA digestion. We compared KAPA and NEB *Taq* polymerases in combination with primer extensions that were progressively more selective on restriction sites, and found greater number of polymorphic SNP loci in pea, white lupin and diploid alfalfa when adopting KAPA with a non-selective primer. This protocol displayed a slight advantage also for tetraploid alfalfa (where SNP calling requires higher read depth). KAPA offered the further advantage of more uniform amplification than NEB over fragment sizes and GC contents. The number of GBS-generated polymorphic markers exceeded 6,500 in two tetraploid alfalfa reference populations and a world collection of lupin genotypes, and 2,000 in different sets of pea or lupin recombinant inbred lines. The predictive ability of GBS-based genomic selection was influenced by the genotype missing data threshold and imputation, as well as by the genomic selection model, with the best model depending on traits and data sets. We devised a simple method for comparing phenotypic vs. genomic selection in terms of predicted yield gain per year for same evaluation costs, whose application to preliminary data for alfalfa and pea in a hypothetical selection scenario for each crop indicated a distinct advantage of genomic selection.

Keywords: GBS, genetic gain, genomic selection, *Lupinus albus*, *Medicago sativa*, *Pisum sativum*, protocol, yield

INTRODUCTION

Next generation sequencing techniques provide many molecular markers at low cost by sequencing single nucleotide polymorphism (SNP) sites in a fraction of the genome. While many array-based procedures require prior knowledge of target sequences, other methods, such as complexity reduction of polymorphic sequences (CRoPS) (van Orsouw et al., 2007) and restriction enzymes

site-associated genomic DNA sequencing (RAD-seq) (Baird et al., 2008), skip sequence discovery and explore SNP variation in DNA fragments cut by a restriction enzyme (RE). The use of methylation sensitive REs, which tend to avoid highly repetitive DNA regions, helps targeting restriction sites that are relatively random and evenly distributed along the genome in gene-rich regions.

In RAD-seq, restriction fragments are ligated to adapters on one end, sheared and size selected, then ligated with adapters on the other end and finally PCR amplified, before sequencing the region flanking the restriction site. Multiplexing many genotypes in a single sequencing reaction can be done using a unique barcode sequence (4 to 8 bp long) in one end of the adapters before ligation. A similar but simplified procedure termed genotyping-by-sequencing (GBS) was proposed by Elshire et al. (2011) using *ApeKI* as a “frequent cutter” RE. GBS skips the shearing and size selection stage, combining the ligation of both adapters into one step. Another improvement of GBS is the modulation of length and nucleotide composition of barcode sequences. The current GBS cost per DNA sample (inclusive of sample preparation) is on the order of 25–30 €. This represents a possible cost reduction of about 40–60% relative to SNP array-based genotyping, with greater reduction for relatively small experiments. At these costs, genomic selection (Heffner et al., 2009) can be applied even to crops of moderate or modest economic importance and/or with no sequenced genome. However, relative to array-based genotyping, GBS presents challenges in coping with missing data and their imputation, and it may require optimization for different species.

The success of GBS depends on the number of polymorphic SNP markers that can be identified. Statistically robust SNP calling depends on the number of sequencing reads per SNP, with a threshold set to 2 or 3 reads for pure lines of inbred species (where heterozygosity is absent), 6 for outbred diploids, and 11 for homozygous loci of an outbred autotetraploid species such as alfalfa, for type I error rate <5%. One way to increase the read depth across all genotypes being sequenced is to minimize the number of fragments that are able to be sequenced. Poland et al. (2012a) developed a modified GBS protocol with double enzyme digestion by *PstI* and *MspI* RE to reduce the number of target sites while increasing their read depth, using a common adapter that allows amplification only of fragments cut by a different RE at each end. Selected amplification may also be pursued by using primers that selectively ligate or amplify a subset of the restriction fragments while using (for example) the *ApeKI* RE (Sonah et al., 2013). GBS protocols that restrict the number of target sites produce markers with greater read depth (for a fixed total number of reads per flow cell) but do not imply necessarily more exploitable SNP markers than the original method by Elshire et al. (2011), because of the lower number of sequenced DNA fragments and, for infrequent cutting RE or RE combinations, because of greater number of large DNA fragments that are amplified less frequently thereby failing to reach the threshold read depth. Finally, the *Taq* polymerase adopted for DNA fragment amplification may change the numbers of successfully sequenced SNP markers, e.g., by using a less selective polymerase

such as KAPA in place of the NEB polymerase adopted in the original method (Elshire et al., 2011).

Greater cultivation of grain and forage legumes is recognized as a key issue for making cropping systems more sustainable in terms of greenhouse gas emissions, energy consumption, soil fertility, and crop diversification (Schneider and Huyghe, 2015), as well as for reducing the marked and increasing insufficiency of feed proteins in large regions such as Europe and China (Pilorgé and Muel, 2016). The main reason for insufficient cultivation of legumes is their modest yielding ability compared with cereals (Reckling et al., 2016), which highlights the importance of exploring the potential of GBS-based genomic selection for higher yield in these crops (Pandey et al., 2016). One study on soybean confirmed the value of genome-enabled predictions by displaying accuracy close to 0.60 (Jarquin et al., 2014). While prediction of pure line performance is the obvious aim of genomic selection in inbred species, predicting the breeding value of candidate parent genotypes for synthetic varieties is the objective of greatest practical interest in outbred species such as alfalfa or other important forage legumes, e.g., white clover or red clover (Annicchiarico et al., 2015a). In a recent study, genomic selection for alfalfa breeding value for forage yield in two contrasting populations achieved an accuracy around 0.35, which could largely offset the gain per unit time from field selection based on progeny test (Annicchiarico et al., 2015b). In another study, accuracies of up to 0.40 were found using a model developed from an initial Cycle 0 population to predict biomass yield of the Cycle 1 population (Li et al., 2015). GBS data may prove valuable also in plant breeding contexts other than genomic selection, namely, in studies of genome-wide association, variety distinctness, diversity, and phylogenetic relationships. Diploid alfalfa such as *Medicago sativa* L. subsp. *caerulea*, while being less useful agronomically than tetraploid alfalfa (subsp. *sativa*), can be studied to produce genomic information of interest also for tetraploid material (Sakiroglu and Brummer, 2016).

With a focus on alfalfa and the cool-season grain legumes pea (*Pisum sativum* L.) and white lupin (*Lupinus albus* L.), this study aimed (i) to assess the effect of RE, *Taq* polymerase and primers on the amount of GBS information generated, and (ii) to report on some successful applications and challenges of genomic selection based on GBS data.

MATERIALS AND METHODS

Experiment 1: Comparison of Restriction Enzyme × *Taq* Polymerase Combinations in *M. sativa* and *M. truncatula*

This study included 2 genotypes of tetraploid *M. sativa* (Altet4 and NECS141) described in Khu et al. (2013), 2 of diploid *M. sativa* (MS-13 and MS-186) described in Han et al. (2012), 2 of diploid *M. sativa* (CC78-68 and CF15-13) described in Li et al. (2011), and 2 reference genotypes of *M. truncatula* [A17, genome-sequenced (Young et al., 2011); and R108].

We compared the utility of two RE protocols [*ApeKI* (Elshire et al., 2011) or *PstI:MspI* (Poland et al., 2012a)] in combination with *Taq* polymerases obtained from either New England Biolabs (NEB) or Kapa Biosystems (KAPA). After DNA extraction, we prepared libraries for each of the 4 RE \times *Taq* combinations (*ApeKI* and NEB, *ApeKI* and KAPA, *PstI:MspI* and NEB, *PstI:MspI* and KAPA). Libraries using *ApeKI* were generated using Elshire et al.'s (2011) protocol with minor modifications. Briefly, 100 ng of each DNA sample (quantified with a Quant-iT PicoGreen dsDNA assay kit, Life Technologies, P7589) was digested with *ApeKI* (NEB, R0643L) and then ligated to a unique barcoded adapter and a common adapter (7.0 ng of the adapter stock were used per the titration test on one alfalfa DNA sample). Equal amounts of the ligated product of each of the eight samples were pooled and cleaned up with QIAquick PCR purification kit (QIAGEN, 28104) for PCR amplification. To generate the *ApeKI* and NEB library, 50 ng template DNA was mixed with NEB 2X *Taq* Master Mix and 2 primers (with 5 nmoles each) in a 50 μ l total volume and amplified on a thermocycler with 18 cycles of 10 s of denaturation at 98°C, 30 s of annealing at 65°C, 30 s extension at 72°C. To generate the *ApeKI* and KAPA library, the only differences were the adoption of the Kapa Library Amplification Readymix (Kapa Biosystems KK2611) instead of NEB and the use of 12 instead of 18 cycles in the amplification program. *PstI:MspI* and NEB and *PstI:MspI* and KAPA were generated according to the protocol by Poland et al. (2012a) with modifications. In each library, we intentionally doubled the amount of *M. truncatula* genotypes, to obtain more reads on those samples.

After de-multiplexing, we identified 64-bp long DNA fragment tags for each genotype using the Stacks pipeline (Catchen et al., 2013), and randomly extracted the same number of reads from each genotype in each library to make fair comparisons among protocols. The same number of reads were randomly extracted using the “fastq-sample” function in the “fastq-tools”¹. Because the lowest number of useful reads for any genotype \times RE \times *Taq* combination was 1.3 M, we extracted 1.25 M reads from each combination to represent approximately the case of 192-plex multiplexing (since the Illumina HiSeq2000 can deliver over 240 M useful reads per lane). We further analyzed data with 2.5 M reads extracted for each genotype, to test approximately the case of 96-plex but excluded the results for tetraploid *M. sativa*, which failed to reach this threshold in all RE \times *Taq* combinations. For each genotype and GBS protocol, we counted the number of tags available for minimum read thresholds of 2, 6, and 11, reporting mean values for the 3 genotype groups (*M. truncatula*; diploid *M. sativa*; tetraploid *M. sativa*).

Experiment 2: Comparisons of KAPA vs. NEB *Taq* Polymerases for Sequencing Bias

We compared KAPA and NEB for selective amplification across DNA fragments that differed for size or for content of

nitrogenous bases, comparing the tag distribution expected from *ApeKI* *in silico* digestion of the *M. truncatula* reference genome with those observed from *ApeKI* digestion with each polymerase (using the highest number of available reads, i.e., 6.6 M). For each tag generated by the two polymerases, we obtained the targeted restriction fragment by BLAST search on the Mt4.0v1 reference genome from genotype A17 downloaded from <http://www.jcvi.org/cgi-bin/medicago/download.cgi>. For DNA fragment size analysis, we computed for KAPA and NEB-based libraries the percentage of tags belonging to each of 15 defined size classes, and compared them with the values expected from *in silico* digestion. For bias relative to content of nitrogenous bases, we computed the percentages of fragments classified to each of seven classes defined based on the GC content of the DNA fragments and compared them with the expected values from *in silico* digestion.

Experiment 3: Comparison of *Taq* Polymerase \times Selective Primer Combinations in Pea, White Lupin, and Alfalfa

This study included the four diploid and two tetraploid genotypes of *M. sativa* described in Experiment 1, six pea genotypes, and four white lupin genotypes. The pea genotypes included three cultivars (Attika, Isard, and Kaspia) that were parents of three connected inbred line populations (Annicchiarico et al., 2017). For each cultivar, we extracted two random genotypes obtained from different commercial seed lots. Although presumably identical, genotyping data revealed genetic differences between the two genotypes of each cultivar in all cases. The lupin germplasm included one genotype from the French cultivar Lucky, and three landrace genotypes that were selected on the basis of their genetic diversity in a prior study performed on a wider genotype set. The source populations of these genotypes were the landraces La568 from Algeria, La646 from the Canary Islands, and LAP123 from Italy, described in the world collection study by Annicchiarico et al. (2010).

The *ApeKI* RE was used for all protocols as described earlier. We compared the 12 protocols represented by KAPA or NEB polymerases in combination with the non-selective primer proposed in the original method (Elshire et al., 2011) or one of 5 3' primers we designed that had 5 to 8 specific bases in order to selectively amplify fragments. All primers were synthesized by Eurofins MWG Operon.

We used the UNEAK pipeline (Lu et al., 2013) for SNP discovery and genotype calling. For fair comparisons, we randomly identified 1.5 M reads from each genotype \times protocol combination in each library as described earlier. A higher number would have resulted in one or more genotypes being dropped from the analysis. For each protocol and genotype combination, we assessed the number of polymorphic SNPs (as 64-bp long sequences with one polymorphism) that were shared by all test genotypes, setting a minimum read depth of 3 for pea and white lupin (inbred species), 6 for diploid alfalfa, and 11 for tetraploid alfalfa. For tetraploid alfalfa, which featured more total reads per

¹<http://homes.cs.washington.edu/~dcjones/fastq-tools/>

genotype, we repeated the assessment for a scenario of 2.5 M total reads per genotype.

Number of Polymorphic Markers and Predictive Ability of Genomic Selection Models in Different Data Sets

This part of the study summarized the number of polymorphic markers and the predictive ability of genomic selection based on cross-validations for yield or quality traits in data sets of tetraploid alfalfa, pea or white lupin. The main sources of data were provided by alfalfa studies in Annicchiarico et al. (2015b) and Biazzi et al. (2017) and the pea study in Annicchiarico et al. (2017), where genotyping protocols, phenotypic procedures for production traits, SNP calling procedures and details of other bioinformatic analyses are reported. Most analyses were performed using various packages of the R software. Some findings from these studies were recalled here to summarize the impact on genomic selection predictive ability of different thresholds for genotype missing data, marker imputation method, and genomic selection model. An initial filtering step excluded markers with a minor allele frequency below 2.5%.

The study by Annicchiarico et al. (2015b) described the genotyping of 154 parent genotypes from a broadly based reference population including Mediterranean germplasm (Me population) and 124 parent genotypes from a broadly based reference population comprising germplasm from the Po Valley, northern Italy (PV population). These germplasm sets were phenotyped separately for forage yield on the basis of densely grown half-sib progenies issued by polycrossing in isolation each set of parents (as convenient for genome-enabled prediction of breeding values: Annicchiarico et al., 2015a). The GBS protocol included *ApeKI* as RE in both populations, while using NEB and KAPA polymerases for PV and Me, respectively. SNP genotype calling distinguished only three classes, namely, the two homozygous ones (AAAA or aaaa), and the heterozygous one (pooling the variants Aaaa, AAaa, and AAAa). A filtering step removed heterozygous loci with less than 4 aligned reads, and homozygous loci with less than 11 reads (thereby reducing the probability to falsely call AAAa or Aaaa heterozygotes as a homozygote to 4.22%).

Biazzi et al.'s (2017) study focused on the same set of half-sib progenies of the Me population, and assessed various quality traits of stems and leaves across three growing conditions (summer harvest, full irrigation; summer harvest, suspended irrigation; autumn harvest). We currently added original genomic selection information on two quality traits, namely, protein content and digestibility of NDF, that were assessed on pooled leaf and stem foliage of the material. For each trait, we compared five genomic selection models, namely, Ridge Regression BLUP, Bayes A, Bayes B, Bayes C, and Bayesian Lasso (Gianola, 2013), for predictive ability based on cross-validations as described in Biazzi et al. (2017) for quality traits of stems and leaves, using a threshold of 30% for missing genotype SNP data and missing data imputation by the *K*-Nearest Neighbor method.

In the pea study (Annicchiarico et al., 2017), 315 F₆ recombinant inbred lines (RILs) belonging to three populations

derived by connected crosses between Attika, Isard, and Kaspia were assessed for grain yield under severe terminal drought stress under a field rainout shelter. The GBS protocol included *ApeKI* as RE and KAPA as *Taq* polymerase. We currently anticipated unpublished genomic selection results for prediction of grain yield in a 3-replicate field experiment carried out in Lodi (northern Italy) under organic farming conditions and autumn sowing in the season 2013–2014. We held a minimum read depth of 4 for SNP genotype calling, because the F₆ generation contained some heterozygous loci. We assessed results for a minimum read depth of 6 as well, obtaining less SNP markers but very similar predictive ability (data not reported), as observed already for grain yield under severe drought (Annicchiarico et al., 2017). Genomic predictions were based on the Ridge Regression BLUP model, 30% threshold for missing genotype SNP data, and missing data imputation by the *K*-Nearest Neighbor method. The model was trained on all populations joined in a single data set and took account of population structure, as performed already in Annicchiarico et al. (2017).

In this paper, predictive ability, i.e., the correlation between genome-based predicted values and observed values, was used as an estimate of prediction accuracy, i.e., the correlation between genome-based predicted values and true breeding values. In several studies on inbred crops, prediction accuracy was estimated by dividing prediction ability by the square root of the broad-sense heritability on a line mean basis, thereby obtaining a higher value that accounts for possible experiment errors in the estimation of breeding values. This correction, however, may introduce a bias when cross-validations are applied to data of the same experiment (Lorenz et al., 2011), as in the current case. We preferred to adopt cautiously lower estimates of prediction accuracy, to minimize the risk of overoptimistic results for genomic selection.

We reported the number of polymorphic markers also for two data sets of white lupin. The former included 288 genotypes sorted out from the world landrace collection in Annicchiarico et al. (2010). The genotypes belonged to seven major historical cropping regions (Madeira-Canaries, Portugal, Spain, Maghreb, Egypt, East Africa, Near East), from each of which we sampled 8 to 10 landrace populations, and 3 or 4 genotypes per landrace population. The latter lupin set included 191 RILs issued by the cross between the cultivar Kiev Mutant and the Ethiopian landrace P27174 [which were used for constructing the first linkage map of this species: Phan et al. (2007)]. RIL DNA samples were obtained from the Institute for Plant Genetics in Poznan in the framework of a joint research work with CREA. GBS protocols for lupin data sets were identical to those for pea that were described in Annicchiarico et al. (2017).

We reported, as a reference, also numbers of polymorphic markers and/or genomic selection predictive ability from other published studies on forage or cool-season grain legumes. The information from the study by Li et al. (2015) was relative to a genomic selection model for forage yield constructed on clonally phenotyped plants when applied to selected intercrossed material evaluated in a further selection stage in the same location. We averaged the results across two target locations (one in NY State and one in Québec) as this scenario, implying predictions

essentially for additive genetic effects, can be compared to prediction of breeding values in the other reported data sets.

Finally, we briefly compared GBS-based genomic selection vs. phenotypic selection in terms of predicted yield gains per unit time in relation to hypothetical selection scenarios and rough estimates of selection costs. Cost estimates (which were inclusive of DNA extraction for GBS) were based on our own experience, recent quotes from genomic platforms, and feedback on phenotyping costs provided by various colleagues and breeding programs, realizing, of course, that GBS costs may decrease in the near future.

RESULTS

Comparison of Restriction Enzyme \times Taq Polymerase Combinations in *M. sativa* and *M. truncatula*

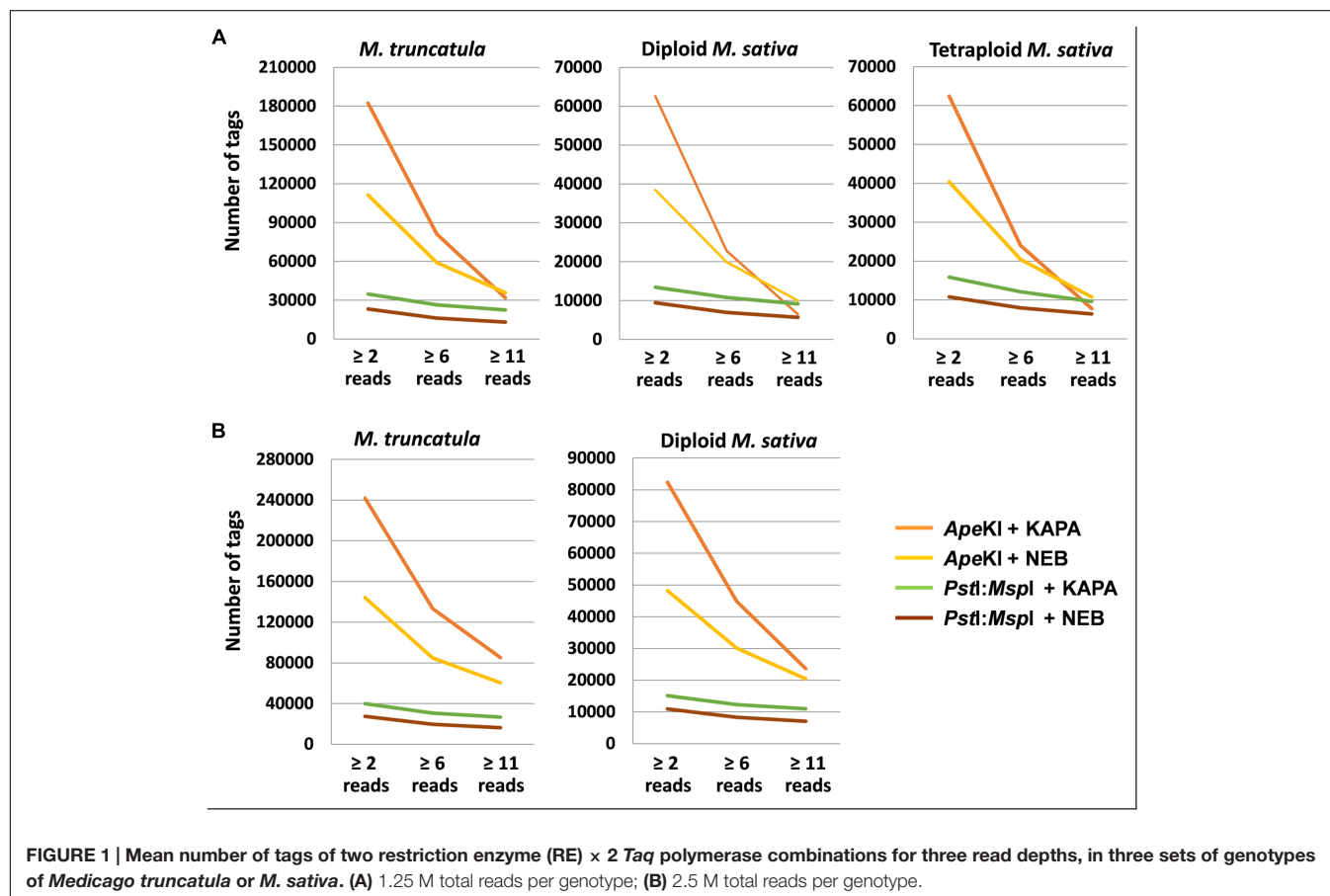
For the scenario of 1.25 M total reads per genotype, the combination of *ApeKI* and KAPA provided a higher number of tags than the other GBS protocols in all sets of genotypes, for minimum read depths of 2 reads per tag (useful for pure lines of inbreds such as *M. truncatula*) or 6 (useful for the outbred diploid *M. sativa* subsp. *sativa*) (Figure 1A). However, the four protocols exhibited few differences and a slight advantage for *ApeKI* and

NEB, when requiring at least 11 reads per tag (as necessary to identify homozygotes for tetraploid alfalfa) (Figure 1A). The responses of diploid and tetraploid alfalfa were nearly identical (Figure 1A), as expected since their 1C genomes are the same size and we held read depth constant across genotypes.

The shift from 1.25 to 2.5 M total reads per genotype improved the relative performance of *ApeKI*-based GBS at higher read depths, particularly when combined with KAPA, whereas *PstI*:*MspI* produced fewer tags than *ApeKI* across all relevant read depths (Figure 1B). Under this scenario, *ApeKI* and KAPA were top-ranking for number of tags even at minimum read depth of 11 (Figure 1B). Results of diploid alfalfa for this read depth are likely to apply also to tetraploid alfalfa under 2.5 M total reads per genotype, when considering the high consistency of results between diploid and tetraploid alfalfa verified under the 1.25 M total read scenario.

Comparisons of KAPA vs. NEB Taq Polymerases for Sequencing Bias

KAPA provided more uniform amplification than NEB over fragment sizes of *M. truncatula*, on the basis of smaller differences between observed and expected tag frequencies for this polymerase. In particular, NEB distinctly overamplified fragments in the range of 100–500 bp (Figure 2), i.e., those that make the greatest contribution to GBS tags (Sonah et al., 2013).



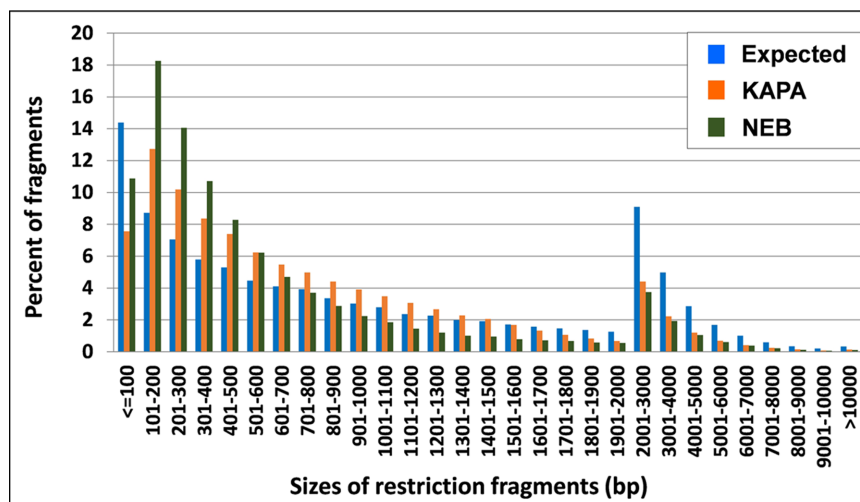


FIGURE 2 | Tag distribution across different sizes of DNA fragments in *M. truncatula* as expected from *ApeKI* *in silico* digestion and observed from 2 *Taq* polymerases.

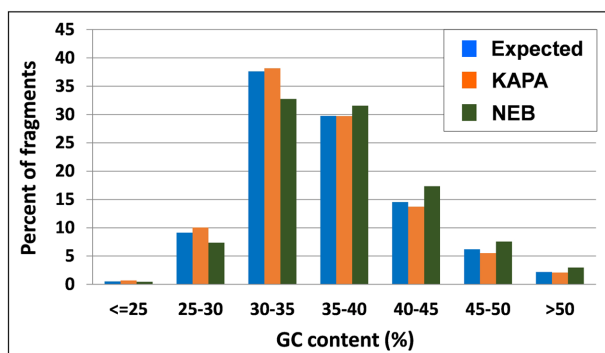


FIGURE 3 | Tag distribution across different GC contents of DNA fragments in *M. truncatula* as expected from *ApeKI* *in silico* digestion and observed from 2 *Taq* polymerases.

The inspection of observed vs. expected frequencies for tag classes that differ for relative content of GC nucleotides revealed more homogeneous amplification by KAPA also across different GC contents. In particular, NEB tended to overamplify the fragments whose GC content exceeded 35%, while underamplifying those with GC content below this level (Figure 3).

Comparison of *Taq* Polymerase by Selective Primer Combinations in Pea, White Lupin, and Alfalfa

The protocol combining KAPA polymerase with the non-selective primer (original method) outperformed any other protocol in terms of number of polymorphic SNP loci for white lupin, pea, and diploid alfalfa (Figure 4). The advantage of this protocol was very large in white lupin and large in pea,

in coincidence with the low minimum read depth required for SNP calling in these inbred species. The advantage was more limited but still sizeable for diploid alfalfa, where it agreed with earlier results for the KAPA vs. NEB comparison in the presence of the non-selective primer that are reported in Figure 1B for the same minimum read depth of 6 and the scenario of 2.5 M total reads per genotype. However, KAPA combined with a selective primer (with minor differences among such primers), or NEB without a selective primer, provided more polymorphic loci than KAPA without a selective primer for the outbred tetraploid *M. sativa* (which requires higher minimum read depth) (Figure 4). In all species, the adoption of a selective primer was more beneficial in combination with KAPA than NEB (Figure 4).

The analysis for the scenario of 2.5 M total reads per genotype, which was performed only for tetraploid alfalfa, indicated a slight advantage of the protocol including KAPA with the non-selective primer (Figure 5), in contrast with results for the scenario of 1.5 M total reads per genotype (Figure 4). Using NEB with the non-selective primer was nearly as good, however (Figure 5).

Number of Polymorphic Markers and Predictive Ability of Genomic Selection Models in Different Data Sets

Genotyping-by-sequencing has been used to genotype several forage or cool-season grain legumes (Table 1). Most reported data sets have been sequenced with approximately 96 samples per lane (in some cases, other samples than those listed in the table were included in a lane in order to generate a 96-plex). All grain legume data sets including a RIL population displayed over 2,300 polymorphic SNP markers, with the exception of one chick pea data set whose lower number of markers may partly be due to more stringent filtering criteria that were adopted for SNP calling or possibly a narrower genetic diversity between the parents of

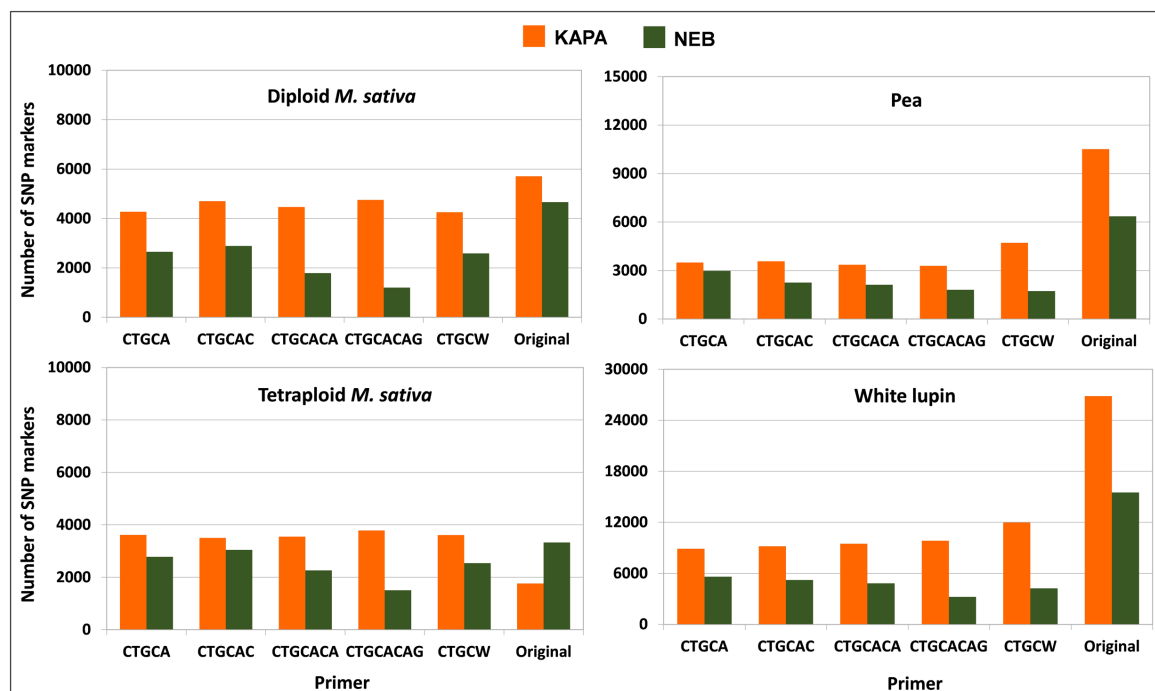


FIGURE 4 | Number of polymorphic SNP markers shared by 6 pea genotypes, 4 white lupin genotypes, 4 genotypes of diploid *M. sativa* and 2 of tetraploid *M. sativa*, for 2 *Taq* polymerase by 6 primer combinations. Minimum read depths of 3 for pea and white lupin, 6 for diploid *M. sativa*, and 11 for tetraploid *M. sativa*; 1.5 M total reads per genotype.

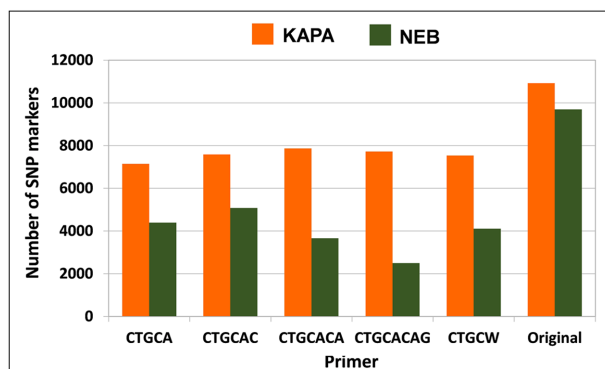


FIGURE 5 | Number of polymorphic SNP markers shared by 2 genotypes of tetraploid *M. sativa*, for 2 *Taq* polymerase by 6 primer combinations. Minimum read depth of 11; 2.5 M total reads per genotype.

the RILs. In white lupin, a world collection of landraces displayed over 2.6-fold more markers than a RIL population, as expected from the greater genetic diversity of this type of germplasm set. Finally, the number of polymorphic SNP markers exceeded 6,500 in all data sets of tetraploid alfalfa.

No comparison between GBS and SNP array procedures for number of polymorphic SNP markers is available for these or other legume data sets. However, the three pairs of parent genotypes that originated the RIL populations of pea displayed,

on average, 3,925 polymorphic SNP markers according to the SNP array facility described by Tayeh et al. (2015a) (Grégoire Aubert and Judith Burstin, pers. comm.). In comparison, the GBS-generated polymorphic SNP markers in the three RIL populations originated by these parents amounted, on average, to 2,547 for the genotype missing data threshold of 30% (Table 1), and 4,409 for the missing data threshold of 50%.

In earlier work of ours on alfalfa and pea (Annicchiarico et al., 2015b, 2017; Biazzi et al., 2017), the predictive ability of genomic selection was affected by the threshold for genotype SNP missing data, the method for imputing missing data, and the genomic selection model. Increasingly relaxed missing data threshold, while increasing the number of polymorphic markers, displayed a peak of predictive ability in the range of 20–40% missing data. This is reported in Figure 6 for forage yield of alfalfa in two data sets, and in Annicchiarico et al. (2017) for grain yield of pea RILs (for minimum read depth of 4 or 6). For missing data imputation, we found an advantage of Random Forest imputation over other methods based on Singular Value Decomposition, Localized Haplotype Clustering, or Mean imputation (Annicchiarico et al., 2015b). However, *K*-Nearest Neighbors imputation proved about as reliable as Random Forest, while being much faster computationally (Nazzicari et al., 2016).

The analysis of various legume data sets indicated that the predictive ability of genomic selection can be affected by the genomic selection model, but no model proved unanimously optimal across different traits or data sets (although in most

TABLE 1 | Mean number of total reads per genotype, and number of polymorphic single nucleotide polymorphism (SNP) markers, in different data sets.

Germplasm set	No. of genotypes	No. of reads/genotype (M)	No. of SNPs	Source
Tetraploid alfalfa, Medit. material ^a	154	2.89	10,339	Annicchiarico et al., 2015b
Tetraploid alfalfa, Po Valley material ^a	124	2.75	6,690	Annicchiarico et al., 2015b
Tetraploid alfalfa, US material ^b	190	2.36	9,906	Li et al., 2015
Pea, Attika × Isard lines ^c	105	2.40	2,386	Annicchiarico et al., 2017
Pea, Kasper × Attika lines ^c	105	1.85	2,506	Annicchiarico et al., 2017
Pea, Kasper × Isard lines ^c	105	2.30	2,750	Annicchiarico et al., 2017
White lupin, Kiev × P27255 lines ^c	191	1.90	2,593	Unpublished data
White lupin, world landrace pool ^c	288	1.66	6,802	Unpublished data
Chickpea, SBD377 × BGD112 lines ^d	95	1.80	3,977	Verma et al., 2015
Chickpea, ICC4958 × ICC1882 lines ^e	210	3.37	828	Jaganathan et al., 2015

^aMinimum no. of reads: 4/11 for heterozygous/homozygous loci; genotype missing data threshold: 30%.

^bMinimum no. of reads: 2/11 for heterozygous/homozygous loci; genotype missing data threshold: 35%.

^cMinimum no. of reads: 4; genotype missing data threshold: 30%.

^dMinimum Qscore of 10 for read retention; unspecified genotype missing data threshold.

^eMinimum Qscore of 20 for read retention; genotype missing data threshold: 50%.

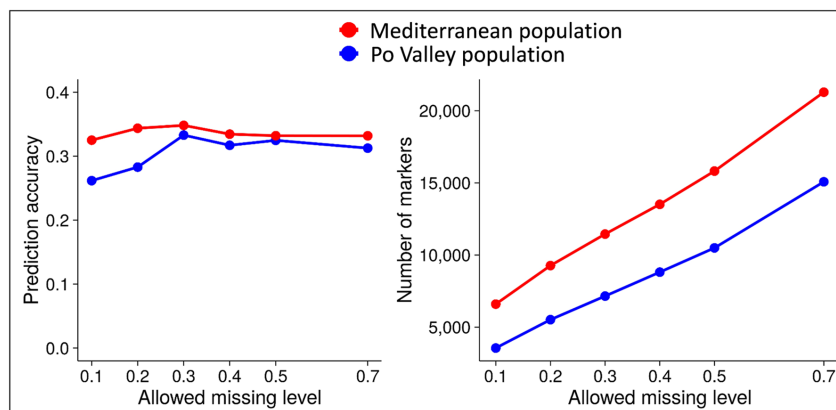


FIGURE 6 | Predictive ability of genomic selection for forage yield and number of polymorphic SNP markers as a function of different genotype missing data thresholds, in two data sets of tetraploid alfalfa. Source of data: Annicchiarico et al. (2015b).

cases, the differences among models were not large). For example, Support Vector Regression with linear kernel outperformed Bayes A, Bayes B, and Bayesian Lasso models for predicting alfalfa forage yield in two data sets (Annicchiarico et al., 2015b), while tending to be outperformed by Bayesian methods (especially Bayesian Lasso) for prediction of pea grain yield (Annicchiarico et al., 2017) and various leaf and stem quality traits of alfalfa (Biazzi et al., 2017). Ridge Regression BLUP (or a model analogous to it, e.g., GBLUP) tended to be among the most accurate models in all of these studies, as well as in two pea studies based on Infinium array SNP markers (Burstin et al., 2015; Tayeh et al., 2015b). In the current comparison of five genomic selection models for two alfalfa quality traits, Ridge regression BLUP and Bayes C revealed a slight advantage for crude protein content and NDF digestibility, respectively, in the absence of marked differences in predicting ability among all tested models (Supplementary Figure 1).

Predictive ability values of the best-performing genomic selection models for yield or key quality traits of alfalfa or pea using GBS-generated SNP data ranged up to 0.72 (Table 2).

Values for alfalfa breeding values ranged between 0.31 and 0.36. Results for pea grain yield were averages of 3 RIL populations reported in Table 1. They were higher than those for alfalfa breeding values for biomass yield, ranging from 0.72 under growing conditions experiencing severe terminal drought to 0.48 for northern Italy under autumn sowing (Table 2). Chick pea results were not available, since data sets in Table 1 were used for GWAS.

Comparison of Genomic vs. Phenotypic Selection Scenarios

With the exception of Li et al. (2015), the predictive ability values reported in Table 2 relate to predictions for the same test environment using cross-validations. Predictions for other test environments (as in the ordinary use of genomic selection) are bound to be less accurate, owing to genotype-by-environment (GE) interactions between the environment(s) used for model definition and those used for application of the model. This is especially true for crop yield, which is usually exposed to wider GE interaction than quality traits. However, the key issue in

TABLE 2 | Predictive ability of genomic selection for genotype breeding value in different data sets.

Germplasm set	Trait	Predictive ability	Source
Tetraploid alfalfa, Medit. material ^a	Biomass yield	0.36	Annicchiarico et al., 2015b
Tetraploid alfalfa, Po Valley material ^a	Biomass yield	0.32	Annicchiarico et al., 2015b
Tetraploid alfalfa, US material ^b	Biomass yield	0.31	Li et al., 2015
Alfalfa, Medit. germplasm ^a	NDF digestibility	0.32	Unpublished data
Alfalfa, Medit. germplasm ^a	Crude protein content	0.32	Unpublished data
Pea, average of three connected crosses ^c	Grain yield, severe drought	0.72	Annicchiarico et al., 2017
Pea, average of three connected crosses ^c	Grain yield, north Italy	0.48	Unpublished data

^aMinimum no. of reads: 4/11 for heterozygous/homozygous loci; genotype missing data threshold: 30%.

^bMinimum no. of reads: 2/11 for heterozygous/homozygous loci; genotype missing data threshold: 35%.

^cMinimum no. of reads: 4; genotype missing data threshold: 30%.

relation to GE interaction for both genomic and phenotypic selection is the ability to predict the genotype breeding values for the target environments of the breeding program. In Li et al. (2015), genomic selection models from NY and Québec were good at predicting each other's phenotyping data, as may be expected from the geographic proximity of their test sites.

Cross-environment predictions can be incorporated into formulas for predicting yield gains from one cycle of phenotypic or genomic selection within a given genetic base. For outbred species such as alfalfa, the predicted gain per year from phenotypic selection (ΔG_P) is:

$$\Delta G_P = (i_P h s_A r_{gP})/t_P$$

where i_P is the standardized selection differential, h is the square root of narrow-sense heritability in the selection conditions, s_A is the standard deviation of breeding values, r_{gP} is the genetic correlation for genotype yield responses between selection and target conditions, and t_P is the number of years for one phenotypic selection cycle. The predicted gain from genomic selection (ΔG_G) is:

$$\Delta G_G = (i_G r_A s_A r_{gG})/t_G$$

where i_G is the standardized selection differential for genomic selection, r_A is the genomic selection accuracy, r_{gG} is the genetic correlation for genotype yield responses between phenotyping conditions for genomic selection modeling and target conditions, and t_G is the duration of one genomic selection cycle. Assuming the same testing conditions ($r_{gP} = r_{gG}$) and selection intensities ($i_P = i_G$), a comparison of phenotypic vs. genomic selection in terms of predicted yield gain per year equates to comparing (h/t_P) vs. (r_A/t_G). For the alfalfa reference population from the Po Valley, the estimated values of $r_A = 0.32$ (Table 1) and $h = 0.46$ [from $h^2 = 0.21$ in Annicchiarico (2015)] suggest that genomic selection would result in higher gain than phenotypic selection if it could halve the duration of one selection cycle. Actually, considering that $t_G = 1$, and $t_P = 5$ (Table 3) when including the time for recombination of selected material and for phenotypic selection along with 1 year for prior production of half-sib families, this criterion was easily met. Even $r_A = 0.15$ would suffice to grant some advantage to genomic selection over

TABLE 3 | Duration of one selection cycle and indicative cost per evaluated genotype, for hypothetical scenarios of phenotypic and genomic selection for higher yield.

Selection	Selection cycle (years)	Cost per genotype (€) ^a
<i>Cool-season grain legume</i>		
Phenotypic (1 site, 2 years, 3 reps)	2	180–230
Genomic	0.5	32–36
Ratio phenotypic/genomic	4	5.0–7.2
<i>Perennial forage legume</i>		
Progeny test (1 site, 3 years, 3 reps)	5	230–280
Genomic	1	32–36
Ratio phenotypic/genomic	5	6.4–8.7

^aExcluding bioinformatics/data analysis work.

Grain legumes: selection of inbred lines; forage legumes: selection of parents for a synthetic variety.

half-sib progeny-based phenotypic selection according to this criterion.

For inbred species, the reported formulas for estimating expected genetic gains hold true, when h is substituted for by the square root of broad-sense heritability on an entry mean basis (H) under the specific conditions adopted for phenotypic selection:

$$H^2 = s_g^2/(s_g^2 + s_{ge}^2/e + s_e^2/er)$$

where s_g^2 , s_{ge}^2 and s_e^2 are components of variance relative to genotype, GE interaction and pooled experiment error, and e and r are the numbers of test environments and experiment replications, respectively. Extensive multi-environment phenotypic selection (high e values) could rise H near unity, but this is usually prevented by its high cost.

The comparison of selection methods for predicted yield gain could incorporate the evaluation cost per genotype. For example, even the limited multi-environment phenotypic selection scenarios hypothesized in Table 3 result, on average, in about 6-fold greater cost for grain legumes and 7.5-fold greater cost for forage legumes of phenotypic selection relative to genomic selection. Thus, for same total evaluation cost of the two methods, more genotypes could be evaluated by GBS, increasing the selection intensity (i_G). For alfalfa, a comparison of phenotypic vs. genomic selection in terms of predicted

yield gains per year for the same overall costs equates to comparing ($h\ i_P/t_P$) vs. ($r_A\ i_G/t_G$). For example, the phenotypic selection based on progeny-testing of 300 alfalfa genotypes aimed to select 15 parents for a synthetic variety (selected fraction = 5%) implies $i_P = 2.06$ (Falconer, 1989), whereas the genomic selection based on evaluating 2250 genotypes (7.5-fold more than phenotypic selection) that aimed to select 15 parents (selected fraction = 0.66%) implies $i_G = 2.80$. For the alfalfa reference population from the Po Valley and the selection scenarios hypothesized in **Table 3**, genomic selection leads to over 4.7-fold greater predicted yield gain per year than phenotypic selection (from $0.46 \times 2.06/5 = 0.189$ for phenotypic selection vs. $0.32 \times 2.80/1 = 0.896$ for genomic selection). From this perspective, $r_A = 0.15$ (which implies over twofold greater predicted gains for genome-enabled selection) would justify the inclusion of genomic selection in breeding schemes, for the cost and selection cycle scenarios reported in **Table 3**.

For pea in Italian environments, assuming for example the selection of 15 genotypes phenotypically out of 300 or genomically out of sixfold more test genotypes (hence, $i_P = 2.06$, and $i_G = 2.39$) under the scenario in **Table 3** ($t_P = 2$; $t_G = 0.5$), and considering $r_A = 0.48$ (i.e., the lower value for pea in **Table 2**) and a cautiously high estimate of $H = 0.84$ that arises from a multi-environment study in Italy by Annicchiarico and Iannucci (2008) with $e = 7$ (rather than $e = 4$ as in **Table 3**), would result in 2.6-fold greater predicted yield gain per year of genomic selection relative to phenotypic selection. The comparison of phenotypic vs. genomic selection equates here to comparing ($H\ i_P/t_P$) vs. ($r_A\ i_G/t_G$) (i.e., $0.84 \times 2.06/2 = 0.865$ for phenotypic selection vs. $0.48 \times 2.39/0.5 = 2.294$ for genomic selection). An $r_A = 0.36$ would provide a twofold advantage for genomic selection under these circumstances.

DISCUSSION

Our results relative to comparisons of major components of GBS protocols (REs, *Taq* polymerases, primers) cannot be considered conclusive, but they indicated that each of these components may have a large effect on the number of SNP markers generated by GBS. Also, they highlighted the importance of investigating combinations of these components (such as different *Taq* polymerases in combination with different primers or REs), because results for each individual component may vary depending on other components of the GBS protocol.

Reducing the number of target sites through DNA digestion by *PstI:MspI* instead of *ApeKI* was not advantageous for diploid alfalfa or *M. truncatula* at convenient read depths for these species. Results for diploid alfalfa provided indirect evidence for the greater interest of *ApeKI* over *PstI:MspI* even for tetraploid alfalfa at 2.5 M total reads per genotype, which is the ordinary scenario for data sets of this crop (**Table 1**). A contributing reason for the advantage obtained from greater genome complexity reduction by using *PstI:MspI* instead of *ApeKI* in Poland et al. (2012a) could be the larger genome of their target species relative to alfalfa (about 19- and 6-fold larger estimated genome for wheat

and barley, respectively). In addition, the current adoption of KAPA polymerase could amplify the advantage of *ApeKI* over *PstI:MspI* in comparison with NEB polymerase, which was used in Poland et al.'s (2012a) study. *ApeKI* proved preferable to two less frequent-cutting enzymes also for cassava, whose genome size is comparable to alfalfa (Hamblin and Rabbi, 2015).

Selective primers displayed an advantage only in the presence of high minimum read depth and low sequencing effort. In alfalfa (requiring 11 as minimum read depth), selective primers proved advantageous at 1.5 M total reads per genotype but not at 2.5 M total reads. Indeed, one may expect greater advantage from greater mean read depth per SNP obtained via reduction of target sites when adopting low sequencing levels. The advantage of the non-selective primer emerged already at 1.5 M total reads per genotype in diploid alfalfa (minimum read depth of 6), and was particularly large for pea and white lupin (minimum read depth of 2). Our results contrast with those by Sonah et al. (2013) for soybean holding 2 as minimum read depth, where selective primers increased the number of polymorphic SNP markers under scenarios of 1–2 M total reads per genotype. This inconsistency encourages further investigations, also in consideration of the small set of genotypes that provided the basis for the polymorphic SNP assessment in these studies [6 to 2 genotypes here; 2 genotypes in Sonah et al. (2013)].

Results for number of tags or polymorphic markers indicated that KAPA can be preferred to NEB *Taq* polymerase for diploid alfalfa, pea, and white lupin. It is preferable also for tetraploid alfalfa in the ordinary scenario of 2.5 M total reads per genotype. Additionally, our results for *M. truncatula* indicated more uniform amplification over fragment sizes and GC contents of this polymerase relative to NEB. This finding has clear potential for improving GBS-based activities on legumes, since most GBS protocols use NEB. Further comparisons are warranted, however, for other forage or grain legume species.

We assumed a minimum read depth of 2 for pea and white lupin in our assessment of GBS protocols. However, a higher value, such as 4, could conveniently be set for lines that may include some degree of heterozygosity, as we did for the pea RILs that underwent our genomic selection assessment. Anyway, best *Taq* polymerase \times selective primer combinations did not change for these species when considering a minimum read depth of 4 instead of 2 (data not reported).

On the whole, our results support the adoption of a single successful GBS protocol for tetraploid and diploid alfalfa, pea, and white lupin, using *ApeKI* and the non-selective primer as in the original method (Elshire et al., 2011) along with KAPA polymerase. The good overall performance of this protocol might serve as a reference for GBS work in other forage or cool-season legumes that lacks an experimental assessment of protocol components. The value of this protocol ought to be reassessed for tetraploid alfalfa if accurate allele dosage information was desired, which would require about 48x read depth to differentiate the heterozygote classes (Uitdewilligen et al., 2013). We did not consider this scenario, because the large sequencing effort required to obtain thousands of markers with sufficient read depth is currently prevented by its high cost.

We are mainly interested in using GBS for genomic selection to improve crop yield, a complex, highly polygenic trait. No statistical model consistently maximized the predictive ability across different data sets, confirming the scope for exploring different models in genomic selection studies. Ridge Regression BLUP (or its analog GBLUP) ought to be included among the tested models in all cases, on the basis of its currently good performance in different situations and its theoretical suitability for a trait controlled by many loci with small effects (as crop yield is expected to be) (Lorenz et al., 2011). Bayesian Lasso proved to be another well-performing model in most of our analyses. Indeed, these models proved well-performing across a range of plant and animal data sets (de los Campos et al., 2013).

There are two other suggestions for data analysis that descend from our experience. One is imputing missing data by Random Forest [in agreement with results for other species: e.g., Poland et al. (2012b)] or by *K*-Nearest Neighbors, until a sequenced genome will allow for using other methods, e.g., Beagle (Nazzicari et al., 2016). The other is the need for assessing the predictive ability across a range of missing data thresholds. The peak of predictive ability that we found between 20 and 40% missing data, which agrees with other results for soybean (Jarquín et al., 2014) and most results for alfalfa reported by Li et al. (2015), is consistent with the expected trade-off between increased information (more markers) and increased noise (higher imputation errors) that arises from increasing threshold for missing data.

Genotyping-by-sequencing can be used to produce high numbers of polymorphic SNP markers for forage and cool-season grain legumes. For three pea RIL populations, the number of polymorphic markers generated by GBS was comparable to that expected at much higher cost by a SNP array genotyping. Particularly when used with best model configurations, GBS-based genomic predictions were sufficiently high for cost-effective exploitation by breeding programs. The occurrence of more accurate predictions in pea than alfalfa could be expected, owing to the much longer linkage disequilibrium and the possibility to thoroughly exploit also non-additive genetic variation that feature the RILs of an inbred crop compared with a set of progeny-tested alfalfa parents. Genomic selection accuracy for soybean grain yield based on cross-validations, which achieved 0.64 (Jarquín et al., 2014), is intermediate between the values of 0.72 and 0.48 that we found for pea grain yield. The difference in prediction accuracy between the two data sets of pea could be attributed to the different ecological complexity of the two phenotyping environments. The environment prone to severe terminal drought, which reproduced a climatically unfavorable Mediterranean environment, was ecologically simpler, because higher genotype yield was strictly associated genetically with an early phenology (Annicchiarico et al., 2017). Higher yield in the autumn-sown, subcontinental-climate environment of northern Italy required genotype adaptation to both low winter temperatures and terminal drought.

We proposed a simple general framework for comparing genomic vs. phenotypic selection in terms of expected genetic

gain, which takes account of differences in selection cycle duration and genotype evaluation cost between the selection methods. These differences can be substantial, and their impact on the relative efficiency of selection methods can be important. We chose only one example scenario among many possible ones for comparing genomic vs. phenotypic selection in each crop, and we lacked estimates for important parameters such as r_{GP} and r_{GG} . In the absence of these estimates, the ability of genomic and phenotypic selection to predict genotype yields in cropping environments other than the test one could be highly informative, but also this information was not available. Though limited, our preliminary comparison revealed a large predicted advantage of genomic selection over phenotypic selection that is encouraging for legume breeding and supports further and more conclusive assessments of genome-enabled predicting ability across a wider set of cropping environments.

Rajsic et al. (2016) proposed another method for comparing genomic vs. phenotypic selection in inbred crops that accounts for different selection costs, in which genomic prediction accuracy is estimated as a function of trait heritability, effective number of chromosome segments underlying the trait, and training population size. When setting $H^2 = 0.85$ for pea, already a twofold cost of phenotypic selection relative to genomic selection would imply some predicted advantage for genomic selection across a wide range of effective number of chromosome segments. With no account for different selection costs or selection cycle duration, simulation results by Viana et al. (2016) for outbred species suggest greater predicted yield gain per selection cycle by genomic selection when h^2 is below 0.30 (as here for alfalfa), for a scenario of 200 genotyped individuals and moderate sequencing effort.

Another issue of interest is the ability of a model set up for a given genetic base to predict the same trait in a different genetic base. Cross-population predictions for alfalfa biomass of Mediterranean germplasm based on a model from Po Valley germplasm or vice versa implied just a moderate loss of predictive ability (25–30%) relative to intra-population predictions (Annicchiarico et al., 2015b). Preliminary results for pea revealed small to quite large loss of predictive ability passing from intra-population to cross-population prediction of grain yield, depending on the pair of RIL populations (Annicchiarico et al., 2017).

We valued genomic selection mainly for its ability to increase the rate of genetic yield gain through shorter selection cycles and more evaluated genotypes for same overall cost. Another contribution of genomic selection to crop improvement could be its unprecedented potential for selecting simultaneously for several traits. This can be particularly important for perennial forage legume breeding, which is constrained by high phenotyping costs and requires at least 10–15 selected genotypes as parents of a synthetic variety (Annicchiarico et al., 2016). For example, selecting 10 genotypes for 4 traits at the modest selection rate of 20% for each trait requires a working population of $[10 \times (1/0.20)^4] = 6,250$ individuals, a number that is hardly workable for phenotypic selection in these crops (particularly when involving a time- and resource-consuming trait, such as forage yield across several harvests and production cycles)

while being within reach for genomic selection (particularly in the perspective of continuously decreasing genotyping costs). The moderately high genome-enabled predictive ability that emerged for two important alfalfa forage quality traits, namely, digestibility of NDF and protein content, has practical interest and supports this perspective use of genomic selection. While higher protein content is beneficial to decrease the dependency of crop-livestock systems from expensive extra-farm feed protein sources, higher NDF digestibility is the main determinant of cattle dry-matter intake and milk yield (Oba and Allen, 1999).

We expect a steep rise in genomic selection studies in forage and cool-season grain legumes in the next few years, especially because of the promising results that have emerged from the first studies (such as those reported here). GBS data will probably be pivotal in this context, owing to their low cost and possible usefulness also for identifying candidate genes in GWAS as soon as a sequenced genome becomes available for these crops. Challenges arising from GBS-based genotyping (adopted protocol; SNP calling procedure; method of missing data imputation; etc.) have not been trivial in pioneering work, but are bound to be overcome by increasing scientific knowledge, availability of sequencing platforms and development of bioinformatic tools.

AUTHOR CONTRIBUTIONS

PA and EB designed and supervised the research work, and obtained financial resources. NN was responsible for bioinformatics analysis. YW was responsible for lab experiments and generation of molecular data. LP and PA were responsible for phenotyping experiments. PA drafted the manuscript. All authors revised, integrated, and approved the manuscript.

REFERENCES

- Annicchiarico, P. (2015). Alfalfa forage yield and leaf/stem ratio: narrow-sense heritability, genetic correlation, and parent selection procedures. *Euphytica* 205, 409–420. doi: 10.1007/s10681-015-1399-y
- Annicchiarico, P., Barrett, B., Brummer, E. C., Julier, B., and Marshall, A. H. (2015a). Achievements and challenges in improving temperate perennial forage legumes. *Crit. Rev. Plant Sci.* 34, 327–380. doi: 10.1080/07352689.2014.898462
- Annicchiarico, P., Boller, B., Brummer, E. C., and Reheul, D. (2016). “Improving the focus of forage breeding research,” in *Breeding in a World of Scarcity*, eds J. Baert, D. Reheul, and I. Roldan (Dordrecht: Springer), 251–269.
- Annicchiarico, P., Harzic, N., and Carroni, A. M. (2010). Adaptation, diversity, and exploitation of global white lupin (*Lupinus albus* L.) landrace genetic resources. *Field Crops Res.* 119, 114–124. doi: 10.1016/j.fcr.2010.06.022
- Annicchiarico, P., and Iannucci, A. (2008). Adaptation strategy, germplasm type and adaptive traits for field pea improvement in Italy based on variety responses across climatically contrasting environments. *Field Crops Res.* 108, 133–142. doi: 10.1016/j.fcr.2008.04.004
- Annicchiarico, P., Nazzicari, N., Li, X., Wei, Y., Pecetti, L., and Brummer, E. C. (2015b). Accuracy of genomic selection for alfalfa biomass yield in different reference populations. *BMC Genomics* 16:1020. doi: 10.1186/s12864-015-2212-y
- Annicchiarico, P., Nazzicari, N., Pecetti, L., Romani, M., Ferrari, B., Wei, Y., et al. (2017). GBS-based genomic selection for pea grain yield under severe terminal drought. *Plant Genome* 10. doi: 10.3835/plantgenome2016.07.0072

FUNDING

This study is part of the ArimNet project REFORMA (‘Resilient, water- and energy-efficient forage and feed crops for Mediterranean agricultural systems’), which generated most of the reported experimental data through funding by the Italian Ministry of Agriculture, Food and Forestry Policy (MiPAAF) and the Samuel Roberts Noble Foundation. Other experimental data were generated by the projects Core Organic II COBRA (‘Coordinating organic plant breeding activities for diversity’) funded by MiPAAF, Qual&Medica (‘High quality alfalfa for the dairy chain’) funded by Fondazione Cassa di Risparmio di Bologna and Regione Emilia-Romagna, and LEGATO (‘Legumes for the agriculture of tomorrow’) funded by the FP7 of the European Commission.

ACKNOWLEDGMENTS

We are grateful to M. Romani, B. Ferrari, A. Tava, S. Proietti, A. Passerini, P. Gaudenzi, and B. Pintus for scientific or technical contribution, M. Książkiewicz for granting our anticipation of some jointly-generated lupin SNP data, and G. Aubert and J. Burstin for providing information on SNP array-based diversity of 3 pea parent genotypes.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fpls.2017.00679/full#supplementary-material>

- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., et al. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE* 3:e3376. doi: 10.1371/journal.pone.0003376
- Biazzi, E., Nazzicari, N., Pecetti, L., Brummer, E. C., Palmonari, A., Tava, A., et al. (2017). Genome-wide association mapping and genomic selection for alfalfa (*Medicago sativa*) forage quality traits. *PLoS ONE* 12:e0169234. doi: 10.1371/journal.pone.0169234
- Burstin, J., Salloignon, P., Chabert-Martinello, M., Magnin-Robert, J.-B., Siol, M., Jacquin, F., et al. (2015). Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics* 16:105. doi: 10.1186/s12864-015-1266-1
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22, 3124–3140. doi: 10.1111/mec.12354
- de los Campos, G., Hickey, J. M., Pong-Wong, R., Daetwyler, H. D., and Calus, M. P. L. (2013). Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193, 327–345. doi: 10.1534/genetics.112.143313
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379. doi: 10.1371/journal.pone.0019379
- Falconer, D. S. (1989). *Introduction to Quantitative Genetics*. Harlow: Longman Scientific & Technical.
- Gianola, D. (2013). Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 194, 573–596. doi: 10.1534/genetics.113.151753

- Hamblin, M. T., and Rabbi, I. Y. (2015). The effects of restriction-enzyme choice on properties of genotyping-by-sequencing libraries: a study in cassava (*Manihot esculenta*). *Crop Sci.* 54, 2603–2608. doi: 10.2135/cropsci2014.02.0160
- Han, Y., Khu, D. M., and Monteros, M. J. (2012). High-resolution melting analysis for SNP genotyping and mapping in tetraploid alfalfa (*Medicago sativa* L.). *Mol. Breed.* 29, 489–501. doi: 10.1007/s11032-011-9566-x
- Heffner, E. L., Sorrells, M. E., and Jannink, J.-L. (2009). Genomic selection for crop improvement. *Crop Sci.* 49, 1–12. doi: 10.2135/cropsci2008.08.0512
- Jaganathan, D., Thudi, M., Kale, S., Azam, S., Roorkiwal, M., Gaur, P. M., et al. (2015). Genotyping-by-sequencing based intra-specific genetic map refines a "QTL-hotspot" region for drought tolerance in chickpea. *Mol. Genet. Genomics* 290, 559–571. doi: 10.1007/s00438-014-0932-3
- Jarquín, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15:740. doi: 10.1186/1471-2164-15-740
- Khu, D., Reyno, R., Han, Y., Zhao, P. X., Bouton, J. H., Brummer, E. C., et al. (2013). Identification of aluminum tolerance quantitative trait loci in tetraploid alfalfa. *Crop Sci.* 53, 148–163. doi: 10.2135/cropsci2012.03.0181
- Li, X., Wang, X., Wei, Y., and Brummer, E. C. (2011). Prevalence of segregation distortion in diploid alfalfa and its implications for genetics and breeding applications. *Theor. Appl. Genet.* 123, 667–679. doi: 10.1007/s00122-011-1617-5
- Li, X., Wei, Y., Acharya, A., Hansen, J. L., Crawford, J. L., Viands, D. R., et al. (2015). Genomic prediction of biomass yield in two selection cycles of a tetraploid alfalfa breeding population. *Plant Genome* 8. doi: 10.3835/plantgenome2014.12.0090
- Lorenz, A. J., Chao, S., Asoro, F. G., Heffner, E. L., Hayashi, T., Iwata, H., et al. (2011). Genomic selection in plant breeding: knowledge and prospects. *Adv. Agron.* 110, 77–123. doi: 10.1016/B978-0-12-385531-2.00002-5
- Lu, F., Lipka, A. E., Glaubitz, J., Elshire, R., Cherney, J. H., Casler, M. D., et al. (2013). Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet.* 9:e1003215. doi: 10.1371/journal.pgen.1003215
- Nazzicari, N., Biscarini, F., Cozzi, P., Brummer, E. C., and Annicchiarico, P. (2016). Marker imputation efficiency for genotyping-by-sequencing data in rice (*Oryza sativa*) and alfalfa (*Medicago sativa*). *Mol. Breed.* 36, 691–16. doi: 10.1007/s11032-016-0490-y
- Oba, M., and Allen, M. S. (1999). Evaluation of the importance of the digestibility of neutral detergent fiber from forage: effects on dry matter intake and milk yield of dairy cows. *J. Dairy Sci.* 82, 589–596. doi: 10.3168/jds.S0022-0302(99)75271-9
- Pandey, M. K., Roorkiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., et al. (2016). Emerging genomic tools for legume breeding: current status and future prospects. *Front. Plant Sci.* 7:455. doi: 10.3389/fpls.2016.00455
- Phan, H. T., Ellwood, S. R., Adhikari, K., Nelson, M. N., and Oliver, R. P. (2007). The first genetic and comparative map of white lupin (*Lupinus albus* L.): identification of QTLs for anthracnose resistance and flowering time and a locus for alkaloid content. *DNA Res.* 14, 59–70. doi: 10.1093/dnares/dsm009
- Pilorgé, E., and Muel, F. (2016). What vegetable oils and proteins for 2030? Would the protein fraction be the future of oil and protein crops? *OCL* 23:D402. doi: 10.1039/c3fo60717h
- Poland, J. A., Brown, P. J., Sorrells, M. E., and Jannink, J. L. (2012a). Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS ONE* 7:e32253. doi: 10.1371/journal.pone.0032253
- Poland, J. A., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012b). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113. doi: 10.3835/plantgenome2012.06.0006
- Rajic, P., Weersink, A., Navabi, A., and Pauls, K. P. (2016). Economics of genomic selection: the role of prediction accuracy and relative genotyping costs. *Euphytica* 210, 259–276. doi: 10.1007/s10681-016-1716-0
- Reckling, M., Bergkvist, G., Watson, C. A., Stoddard, F. L., Zander, P. M., Walker, R. L., et al. (2016). Trade-offs between economic and environmental impacts of introducing legumes into cropping systems. *Front. Plant Sci.* 7:669. doi: 10.3389/fpls.2016.00669
- Sakiroglu, M., and Brummer, E. C. (2016). Identification of loci controlling forage yield and nutritive value in diploid alfalfa using GBS-GWAS. *Theor. Appl. Genet.* 130, 261–268. doi: 10.1007/s00122-016-2782-3
- Schneider, A., and Huyghe, C. (2015). *Les Légumineuses Pour des Systèmes Agricoles et Alimentaires Durables*. Versailles: Editions Quae.
- Sonah, H., Bastien, M., Iqura, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* 8:e54603. doi: 10.1371/journal.pone.0054603
- Tayeh, N., Aluome, C., Falque, M., Jacquin, F., Klein, A., Chauveau, A., et al. (2015a). Development of two major resources for pea genomics: the GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map. *Plant J.* 84, 1257–1273. doi: 10.1111/tpj.13070
- Tayeh, N., Klein, A., Le Paslier, M.-C., Jacquin, F., Houtin, H., Rond, C., et al. (2015b). Genomic prediction in pea: effect of marker density and training population size and composition on prediction accuracy. *Front. Plant Sci.* 6:941. doi: 10.3389/fpls.2015.00941
- Uitdewilligen, J. G. A. M. L., Wolters, A. M. A., D'hoop, B. B., Borm, T. J. A., Visser, R. G. F., and van Eck, H. J. (2013). A next-generation sequencing method for genotyping-by-sequencing of highly heterozygous autotetraploid potato. *PLoS ONE* 8:e62355. doi: 10.1371/journal.pone.0062355
- van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeijs, S., Verstegen, E., et al. (2007). Complexity reduction of polymorphic sequences (CRoPS): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS ONE* 2:e1172. doi: 10.1371/journal.pone.0001172
- Verma, S., Gupta, S., Bandhiwal, N., Kumar, T., Bharadwaj, C., and Bhatia, S. (2015). High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using Genotyping-by-Sequencing (GBS). *Sci. Rep.* 5:17512. doi: 10.1038/srep17512
- Viana, J. M. S., Piepho, H.-P., and Silva, F. F. (2016). Quantitative genetics theory for genomic selection and efficiency of breeding value prediction in open-pollinated populations. *Sci. Agric.* 73, 243–251. doi: 10.1590/0103-9016-2014-0383
- Young, N. D., Debellé, F., Oldroyd, G. E., Geurts, R., Cannon, S. B., Udvardi, M. K., et al. (2011). The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524. doi: 10.1038/nature10625

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Annicchiarico, Nazzicari, Wei, Pecetti and Brummer. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

[@frontiersin](https://twitter.com/frontiersin)



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership