

Creation and utilization of crop germplasm resources

Edited by

Chuanzhi Zhao, Rajeev K. Varshney and Xia Xin

Published in

Frontiers in Plant Science



FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83251-630-0
DOI 10.3389/978-2-83251-630-0

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: frontiersin.org/about/contact

Creation and utilization of crop germplasm resources

Topic editors

Chuanzhi Zhao — Shandong Academy of Agricultural Sciences, China

Rajeev K. Varshney — Murdoch University, Australia

Xia Xin — Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, China

Topic coordinator

Sunil S. Gangurde — University of Georgia, United States

Citation

Zhao, C., Varshney, R. K., Xin, X., eds. (2023). *Creation and utilization of crop germplasm resources*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83251-630-0

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.

Table of contents

- 05 **Editorial: Creation and utilization of crop germplasm resources**
Chuanzhi Zhao, Sunil S. Gangurde, Xia Xin and Rajeev K. Varshney
- 09 **Combining Genome-Wide Association Study and Gene-Based Haplotype Analysis to Identify Candidate Genes for Alkali Tolerance at the Germination Stage in Rice**
Song Mei, Guogen Zhang, Jing Jiang, Jingbing Lu and Fan Zhang
- 22 **Development and Application of Intragenic Markers for 14 Nitrogen-Use Efficiency Genes in Rice (*Oryza sativa* L.)**
Pingbo Li, Zhen Li, Xu Liu, Hua Zhang, Qingguo Wang, Nana Li, Hanfeng Ding and Fangyin Yao
- 33 **Small RNA and Degradome Deep Sequencing Reveals the Roles of microRNAs in Peanut (*Arachis hypogaea* L.) Cold Response**
Xin Zhang, Chao Ren, Yunyun Xue, Yuexia Tian, Huiqi Zhang, Na Li, Cong Sheng, Huifang Jiang and Dongmei Bai
- 48 **Selection Signatures in Chinese Sorghum Reveals Its Unique Liquor-Making Properties**
Liyi Zhang, Yanqing Ding, Jianxia Xu, Xu Gao, Ning Cao, Kuiying Li, Zhou Feng, Bing Cheng, Lengbo Zhou, Mingjian Ren, Xiaochun Lu, Zhigui Bao, Yuezhi Tao, Zhanguo Xin and Guihua Zou
- 61 **Artificial Selection Trend of Wheat Varieties Released in Huang-Huai-Hai Region in China Evaluated Using DUS Testing Characteristics**
Liyuan Wang, Yongsheng Zheng, Lili Duan, Mumu Wang, Hui Wang, Hua Li, Ruyu Li and Han Zhang
- 72 **Genomic Analysis Provides Insights Into the Plant Architecture Variations in *in situ* Conserved Chinese Wild Rice (*Oryza rufipogon* Griff.)**
Ziyi Yang, Yilin Zhang, Meng Xing, Xiaowen Wang, Zhijian Xu, Jingfen Huang, Yanyan Wang, Fei Li, Yamin Nie, Jinyue Ge, Danjing Lou, Ziran Liu, Zhenyun Han, Yuntao Liang, Xiaoming Zheng, Qingwen Yang, Hang He and Weihua Qiao
- 81 **Germplasm Sources, Genetic Richness, and Population Differentiation of Modern Chinese Soybean Cultivars Based on Pedigree Integrated With Genomic-Marker Analysis**
Chunyan Li, Wubin Wang, Yongpeng Pan, Fangdong Liu, Jianbo He, Chuanxiang Liu, Jiqiu Cao, Xiaoyan Zhang, Jinming Zhao and Junyi Gai
- 97 **GWAS and RNA-seq analysis uncover candidate genes associated with alkaline stress tolerance in maize (*Zea mays* L.) seedlings**
Chunxiang Li, Yue Jia, Runyu Zhou, Liwei Liu, Mengna Cao, Yu Zhou, Zhenhua Wang and Hong Di
- 113 **Overexpression of *ZmIPT2* gene delays leaf senescence and improves grain yield in maize**
Yongfeng Song, Chunxiang Li, Yong Zhu, Pei Guo, Qi Wang, Lin Zhang, Zhenhua Wang and Hong Di

- 129 **Molecular cytogenetic characterization of a new wheat-*Thinopyrum intermedium* homoeologous group-6 chromosome disomic substitution line with resistance to leaf rust and stripe rust**
Xiaojun Zhang, Jianbo Li, Yudi Ge, Haixia Guan, Guangrong Li, Shuwei Zhang, Xiaolu Wang, Xin Li, Zhijian Chang, Peng Zhang, Juqing Jia and Cheng Liu
- 138 **Optimized methods for random and targeted mutagenesis in field pea (*Pisum sativum* L.)**
Prashant Kumar Pandey, Pankaj Bhowmik and Sateesh Kagale
- 151 **Co-overexpression of *AtSAT1* and *EcPAPR* improves seed nutritional value in maize**
Xiaoli Xiang, Binhua Hu, Zhigang Pu, Lanying Wang, Thomas Leustek and Changsheng Li
- 169 **Targeted metabolome analysis reveals accumulation of metabolites in testa of four peanut germplasms**
Kun Zhang, Jing Ma, Sunil S. Gangurde, Lei Hou, Han Xia, Nana Li, Jiaowen Pan, Ruizheng Tian, Huailing Huang, Xingjun Wang, Yindong Zhang and Chuanzhi Zhao
- 183 **Isolation and identification of *SiCOL5*, which is involved in photoperiod response, based on the quantitative trait locus mapping of *Setaria italica***
Fei-fei Li, Jia-hong Niu, Xiao Yu, Qing-hua Kong, Run-feng Wang, Ling Qin, Er-ying Chen, Yan-bing Yang, Zhen-yu Liu, Li-na Lang, Hua-wen Zhang, Hai-lian Wang and Yan-an Guan
- 197 **Integrated analysis of carotenoid metabolites and transcriptome identifies key genes controlling carotenoid compositions and content in sweetpotato tuberous roots (*Ipomoea batatas* L.)**
Ruixue Jia, Rong Zhang, Sunil S. Gangurde, Chaochen Tang, Bingzhi Jiang, Guilan Li and Zhangying Wang
- 211 **Assessing the heat sensitivity of Urdbean (*Vigna mungo* L. Hepper) genotypes involving physiological, reproductive and yield traits under field and controlled environment**
Shikha Chaudhary, Uday Chand Jha, Pronob J. Paul, P. V. Vara Prasad, Kamal Dev Sharma, Sanjeev Kumar, Debjyoti Sen Gupta, Parul Sharma, Sarvjeet Singh, Kadambot H. M. Siddique and Harsh Nayyar
- 229 **Whole genome resequencing identifies candidate genes and allelic diagnostic markers for resistance to *Ralstonia solanacearum* infection in cultivated peanut (*Arachis hypogaea* L.)**
Chong Zhang, Wenping Xie, Huiwen Fu, Yuting Chen, Hua Chen, Tiecheng Cai, Qiang Yang, Yuhui Zhuang, Xin Zhong, Kun Chen, Meijia Gao, Fengzhen Liu, Yongshan Wan, Manish K. Pandey, Rajeev K. Varshney and Weijian Zhuang



OPEN ACCESS

EDITED AND REVIEWED BY

Nunzio D'Agostino,
University of Naples Federico II, Italy

*CORRESPONDENCE

Chuanzhi Zhao

✉ chuanzhiz@126.com

Sunil S. Gangurde

✉ sunil.gangurde@uga.edu

SPECIALTY SECTION

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

RECEIVED 08 January 2023

ACCEPTED 17 January 2023

PUBLISHED 25 January 2023

CITATION

Zhao C, Gangurde SS, Xin X and
Varshney RK (2023) Editorial: Creation and
utilization of crop germplasm resources.
Front. Plant Sci. 14:1140037.
doi: 10.3389/fpls.2023.1140037

COPYRIGHT

© 2023 Zhao, Gangurde, Xin and Varshney.
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(s) 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.

Editorial: Creation and utilization of crop germplasm resources

Chuanzhi Zhao^{1,2*}, Sunil S. Gangurde^{3,4*}, Xia Xin⁵
and Rajeev K. Varshney⁶¹Institute of Crop Germplasm Resources (Institute of Biotechnology), Shandong Academy of Agricultural Sciences; Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan, China, ²College of Life Sciences, Shandong Normal University, Jinan, China, ³Crop Protection and Management Research Unit, United States Department of Agriculture-Agricultural Research Service (USDA-ARS), Tifton, GA, United States, ⁴Department of Plant Pathology, University of Georgia, Tifton, GA, United States, ⁵National Crop GeneBank, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ⁶State Agricultural Biotechnology Centre, Centre for Crop and Food Innovation, Food Futures Institute, Murdoch University, Murdoch, WA, Australia

KEYWORDS

germplasms, mutants, gene identification, QTLs, molecular breeding

Editorial on the Research Topic

Creation and utilization of crop germplasm resources

Introduction

The world's population will almost exceed 8 billion in 2022 (<https://www.unfpa.org/swp2022>). In contrast, the available area of cultivated land is decreasing every year. Therefore, it is a big challenge for international agriculture to ensure the food security of such a huge population (Pathirana and Carimi, 2022). With the recent COVID-2019 pandemic, global warming, economic slowdowns, poverty, and inequality have increased unemployment, resulted in food insecurity, and affected millions of people globally (FAO, 2021). The State of Food Security and Nutrition in the World 2021, <https://www.fao.org/3/cb4474en/online/cb4474en.html>. To address the increasing food demands, it is crucial to develop improved crop varieties with enhanced yield through the utilization of diverse germplasm resources (Varshney et al., 2021a).

Crop germplasm resources (CGRs) refer to the genetic material passed from parent to offspring of a crop, which is usually found in a specific variety. Diversity of germplasm is essential for improving multiple traits and increasing genetic gain. Due to the intensive production methods of modern agriculture, many local varieties of crops have been replaced by a few improved elite varieties. This has resulted in decreased genetic diversity in most of the crops. Several cultivated crops lost their important alleles due to selection during human civilization. Crops with a narrow genetic base make it difficult to breed new varieties with high yield and quality; for instance, peanut has resulted in a narrow genetic base during human civilization (Gangurde et al., 2019). Beneficial wild alleles associated with disease resistance and yield traits were lost during domestication and are needed today in modern breeding programs (Bohra et al., 2021; Varshney et al., 2021b). For instance, numerous such wild species from the genera *Aegilops*, *Agropyron*, *Amblyopyrum*, *Dasypyrum*, *Elymus*, *Leymus*, *Pascopyrum*, *Roegneria*, and *Thinopyrum* are growing in a wide range of areas and adapting to diverse climates, and most of them have been used for the genetic improvement of wheat (Soreng et al., 2015; Laugerotte et al., 2022). Similarly, *Haynaldia villosa* is a diploid wild grass with numerous traits similar to those of wheat, including high protein content,

high resistance to cold, and resistance to powdery mildew (Cao et al., 2011). The wheat–*H. villosa* translocation line 6VS/6AL has been applied as the backbone parent of wheat breeding, resulting in the release of dozens of commercial cultivars in China, the USA, and Canada (Cao et al., 2011). As one of the most important oil crops worldwide, cultivated peanut is an allotetraploid, derived from hybridization between the progenitor species *A. duranensis* and *A. ipaensis* (Lavia et al., 2011; Kochert et al., 1996; Halward et al., 1992; Zhao et al., 2021). There are more than 80 species of the genus *Arachis* classified into nine sections, providing great potential to diversify the cultivated gene pool for peanut breeding. One of the main cultivated peanut cultivars, Yuanza9102, showed high resistance to *Ralstonia solanacearum* due to the introgression of the resistance gene of the wild diploid peanut species *Arachis diogeni* (Wang et al., 2018; Han et al., 2022). Therefore, collection, creation, evaluation, conservation, and utilization of germplasm resources, especially from wild species, can help improve the cultivated gene pool.

In 1920, a Soviet geneticist named Nicolai Vavilov realized that most crops had lost their diversity, and he created the first gene bank in Petrograd by doing an extensive series of expeditions worldwide (Janick, 2015). Later, several gene banks were developed in the USA, Western Europe, and Australia (Pathirana and Carimi, 2022). In 1991, the International Plant Genetic Resources Institute (IPGRI) signed an agreement in Rome, Italy, and developed the Crop Genetic Resources Collaboration Network to conserve and utilize germplasm resources. The network could collect ~200,000 accessions from 136 countries (Pathirana and Carimi, 2022). Utilization of these germplasm resources to enrich the cultivated crop varieties is a burning task for current breeding programs. Advances in genomics, phenomics, and bioinformatics have opened new frontiers for the conservation and utilization of germplasm resources. In our Research Topic entitled “creation and utilization of crop germplasm resources,” we published 17 research articles that shed light on the genetic diversity and evolution of crop germplasm resources, including wild relatives, landraces, and cultivated varieties, by using whole genome resequencing and association analysis, phenotyping, and Distinctness, Uniformity, and Stability (DUS) characterization, concerning ideotypes, tolerance to biotic/abiotic stresses, high yield, and nutritional traits. Furthermore, the included articles contributed to the identification of important genes/QTLs and uncovered the molecular mechanisms of those important traits.

Germplasm characterization, genetic relationship, and population structure of collected germplasms

Germplasm characterization is very important to identify potential sources of traits of interest. For instance, 560 soybean cultivars, comprised of 279 cultivars from Northeast and Northwest China (NNC), 155 cultivars from the Huang-Huai-Hai Valleys (HHH), and 126 from Southern China (SC), were collected and designated as Modern Chinese Soybean Cultivars (MCSCs). Population structure analysis identified 13 maturity groups in soybeans in different ecoregions. The germplasms of NNC showed high allele diversity but were distant when compared with those of HHH and SC. Eleven major core-terminal ancestor-derived families, including four derived from ancestors in NNC,

four from HHH, and three from SC, contain 463 (82.68%) MCSCs with some cross-distribution among ecoregions (Li et al.). Based on these findings, we can predict how important it is to exchange germplasm to enhance the genetic potential of crops.

Distinctiveness, uniformity, and stability (DUS) are the criteria for a new variety release for cultivation in a farmer's field. A total of 195 wheat varieties from the Huang-Huai-Hai region of China were evaluated for 35 DUS characteristics. Of the 35 traits, eight characteristics varied significantly, with the most variation in the sterile spikelet. Artificial selection trends in flag leaf length and flag leaf width, grain number per ear, and grain volume weight showed an overall upward trend. In contrast, plant height showed a downward trend. These findings indicated that selection of some non-economic characteristics of wheat varieties, such as awn color, stem color, and glume color, seemed to be able to enrich the genetic diversity of varieties in the Huang-Huai-Hai region (Wang et al.).

The brewing industry in China is very popular, and Chinese sorghum has been an important ingredient in brewing since ancient times. Re-sequencing of 244 Chinese sorghum accessions resulted in three genetic clusters, namely the Northern, the Southern, and the Chishui groups. An important selective sweep region was identified with homologous genes involved in grain size, pericarp thickness, and the architecture of the inflorescence. These results also suggested that pericarp strength determines the ability of grain to resist repeated cooking during the brewing process (Zhang et al.).

The alleles from *in situ* conserved wild rice (*Oryza rufipogon* Griff.), also called “Guiping wild rice,” can improve rice production worldwide. A comparison of resequencing data between Guiping wild rice populations and *O. rufipogon* and *Oryza sativa* populations indicated that the *in situ* conserved wild rice population has a unique genetic structure, with genes that were introgressed from aromatic and *O. sativa* ssp. *indica* and *japonica* populations (Yang et al.).

Genome-wide association study and multiple omics facilitated mining of beneficial genes/loci

The comparative metabolomics study of four peanut testa colors, including pink, purple, red, and white, identified 78 differentially accumulated metabolites. The proanthocyanidins are most abundant in pink testa, whereas, the red testa accumulated more isoflavones, flavonols, and anthocyanidins (Zhang et al.). Similarly, the integrated analysis of the metabolome and transcriptome identified 14 carotenoids in five sweet potato cultivars and 27 differentially expressed genes involved in carotenoid metabolism, respectively (Jia et al.). To identify candidate genes and allelic diagnostic markers for resistance to *R. solanacearum* of cultivated peanut, QTL-seq analysis identified the resistance gene mapped in a 7.2 Mb physical region of chromosome 12 of *Arachis hypogaea*, and eight nucleotide binding site leucine rich repeat genes were highlighted. Interestingly, two diagnostic SNP markers were developed and validated for breeding disease-resistant peanut varieties (Zhang et al.). Cold stress is an important abiotic stress affecting plant growth and development by interfering with physiological processes. A combinatorial approach to small RNA and degradome sequencing identified 407 known miRNAs and 143 novel

peanut-specific miRNAs. Transient expression analysis in *Nicotiana benthamiana* showed that miR160, miR482, and miR2118 may play positive roles and that miR396, miR162, and miR1511 play negative roles in the regulation of peanut cold tolerance (Zhang et al.).

In this special issue, we also reported genetic dissection and candidate gene/locus identification for several abiotic stresses, such as salinity tolerance, nitrogen utilization efficiency, cold tolerance, and photoperiod response. Salinity–alkalinity is among the serious abiotic stresses limiting the yield potential of several crops. For instance, an association panel of 200 maize lines identified nine SNPs and associated candidate genes with alkaline tolerance in maize seedlings. RNA-Seq analysis confirmed five hub genes involved in alkaline tolerance (Li et al.). The salinity–alkalinity tolerance of rice varieties at the germination stage is very important because of the widespread use of direct seeding technology in rice. Based on the evaluation of seven germination-related traits on a 428-rice diversity panel, Xian/indica accessions showed generally higher tolerance to alkali stress than Geng/japonica accessions. Further, the GWAS identified 90 loci and eight candidate genes associated with the alkali tolerance. Interestingly, a negative regulator of alkali tolerance gene LOC_Os09g25060 (OsWRKY76) was identified (Mei et al.). Among two subspecies of rice, geng/japonica has significantly lower nitrogen-use efficiency (NUE) than xian/indica. Haplotype analysis based on 14 cloned genes of NUE and 36 rice germplasm lines developed 18 intragenic markers. The results reported 12 NUE genes, which are mostly present in XI accessions. The elite haplotype of gene *DEP1* is fixed in geng/japonica cultivars, and elite haplotypes of genes *MYB61* and *NGR5* have been introduced into some approved geng/japonica cultivars (Li et al.). Though foxtail millet is a model crop, it is not globally distributed and utilized due to its photoperiod sensitivity. Genetic mapping for photoperiod sensitivity using a RIL population (Longgu 3 × Canggu 3) identified 21 QTLs and 116 candidate genes. A candidate gene, *SiCOL5*, was identified as photoperiod-sensitive and regulated by biological rhythm-related genes (Li et al.), which might shed light on photoperiod-tolerant breeding.

Identification of the function of candidate genes

Maize seeds are deficient in the essential amino acids cysteine (Cys) and methionine (Met). The improved highest Met maize line, pRbcS : AtSAT1-pRbcS : EcPAPR, increased Met by 2.24-fold. But the plants of pRbcS : AtSAT1-pRbcS : EcPAPR showed progressively severe defects in plant growth, including early senescence, stunting, and dwarfing, indicating that excessive sulfur assimilation has an adverse effect on plant development. The transcriptome analysis of maize leaves using pRbcS : AtSAT1-pRbcS : EcPAPR identified 3,274 differentially expressed genes associated with Met homeostasis. Two genes, serine/threonine-protein kinase (CCR3) and heat shock 70 kDa protein (HSP), were identified in the core of the leaves and endosperm, respectively (Xiang et al.). Isopentenyl transferase (IPT) is an important rate-limiting enzyme in cytokinin (CTK) synthesis in plants. In maize, over-expression of *ZmIPT2* led to delayed senescence of leaves and 17.71%–20.29% increases in grain yield, providing new insight for the breeding of new high-yield transgenic maize varieties (Song et al.).

Creation and utilization of new germplasms by mutagenesis and introgression of foreign DNAs or chromosome fragments

Mutation breeding can help expand the genetic base by producing novel alleles with the help of chemical and physical mutagens. In field peas, physical and chemical mutagenesis were carried out using gamma irradiation and ethyl methanesulfonate (EMS), respectively. A gamma radiation dose of 225 Gy and an EMS concentration of 5 mM were selected as optimal dosages for mutagenesis in field peas. PEG-mediated transformation and gene editing of the *LOX* gene were carried out using the CRISPR/Cas system, providing the platform for creating new germplasms (Pandey et al.). *Thinopyrum intermedium* (JjJJsStSt, 2n = 6x = 42), with good resistance to many wheat diseases, is known as one of the most important closely related wild species of wheat. A new line, CH51, was developed from the BC₁F₈ progeny of a partial wheat-*T. intermedium* amphiploid TAI8335 and wheat cultivar (cv.) Jintai 170. The CH51 showed high levels of resistance to the prevalent Chinese leaf rust and stripe rust races in the field and can be used to increase the resistance of wheat (Zhang et al.).

Author contributions

CZ and SG drafted the manuscript. XX and RV provided input and comments to the draft. All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This research is supported by the National Key Research and Development Program of China (2022YFD1200401), the Agricultural Science and Technology Innovation Program (CAAS-ASTIP-2021-ICS01), the Key Research and Development Project of Shandong Province (2020LZGC001, 2021LZGC025, and 2022LZGC022), and the Agricultural Scientific and Technological Innovation Project of the Shandong Academy of Agricultural Sciences.

Acknowledgments

We would like to thank all the contributing authors for their interest in our Research Topic. We also thank all the reviewers for their valuable comments and suggestions for each article. We gratefully thank Prof. Xingjun Wang and Dr. Dong Wang from the Shandong Academy of Agricultural Sciences for providing a proposal for the Research Topic.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S. R., et al. (2021). Reap the crop wild relatives for breeding future crops. *Trends Biotechnol.* 40 (4):412–431. doi: 10.1016/j.tibtech.2021.08.009
- Cao, A., Xing, L., Wang, X., Yang, X., Wang, W., Sun, Y., et al. (2011). Serine/threonine kinase gene stpK-V, a key member of powdery mildew resistance gene Pm21, confers powdery mildew resistance in wheat. *Proc. Natl. Acad. Sci.* 108 (19), 7727–7732. doi: 10.1073/pnas.1016981108
- FAO. (2021). *The State of Food Security and Nutrition in the World*. Available at: <https://www.fao.org/3/cb4474en/online/cb4474en.html>.
- Gangurde, S. S., Kumar, R., Pandey, A. K., Burow, M., Laza, H. E., Nayak, S. N., et al. (2019). Climate-smart groundnuts for achieving high productivity and improved quality: Current status, challenges, and opportunities. *Genomic designing climate-smart oilseed crops*. 2019, 133–172. doi: 10.1007/978-3-319-93536-2_3
- Halward, T., Stalker, T., LaRue, E., and Kochert, G. (1992). Use of single-primer DNA amplifications in genetic studies of peanut (*Arachis hypogaea* L.). *Plant Mol. Biol.* 18 (2), 315–325. doi: 10.1007/BF00034958
- Han, S., Zhou, X., Shi, L., Zhang, H., Geng, Y., Fang, Y., et al. (2022). AhNPR3 regulates the expression of WRKY and PR genes, and mediates the immune response of the peanut (*Arachis hypogaea* L.). *Plant J.* 110 (3), 735–747. doi: 10.1111/tpj.15700
- Janick, J. (2015). Nikolai Ivanovich vavilov: Plant geographer, geneticist, martyr of science. *HortScience* 50 (6), 772–776. doi: 10.21273/HORTSCI.50.6.772
- Kochert, G., Stalker, H. T., Gimenes, M., Galgano, L., Lopes, C. R., and Moore, K. (1996). RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *arachis hypogaea* (Leguminosae). *Am. J. Bot.* 83 (10), 1282–1291. doi: 10.1002/j.1537-2197.1996.tb13912.x
- Laugerotte, J., Baumann, U., and Sourdille, P. (2022). Genetic control of compatibility in crosses between wheat and its wild or cultivated relatives. *Plant Biotechnol. J.* 20(5):812–832. doi: 10.1111/pbi.13784
- Lavia, G. I., Ortiz, A. M., Robledo, G., Fernández, A., and Seijo, G. (2011). Origin of triploid arachis pintoi (Leguminosae) by autopolyploidy evidenced by FISH and meiotic behaviour. *Ann. Bot.* 108 (1), 103–111. doi: 10.1093/aob/mcr108
- Pathirana, R., and Carimi, F. (2022). Management and utilization of plant genetic resources for a sustainable agriculture. *Plants* 11 (15), 2038. doi: 10.3390/plants11152038
- Soreng, R., Peterson, P., Romaschenko, K., Davidse, G., Zuloaga, F., Judziewicz, E., et al. (2015). A worldwide phylogenetic classification of the poaceae (Gramineae). *J. Syst. Evol.* 53, 117–137. doi: 10.1111/jse.12150
- Varshney, R. K., Bohra, A., Roorkiwal, M., Barmukh, R., Cowling, W. A., Chitkineni, A., et al. (2021a). Fast-forward breeding for a food-secure world. *Trends Genet.* 37 (12), 1124–1136. doi: 10.1016/j.tig.2021.08.002
- Varshney, R. K., Bohra, A., Yu, J., Graner, A., Zhang, Q., and Sorrells, M. E. (2021b). Designing future crops: Genomics-assisted breeding comes of age. *Trends Plant Sci.* 26 (6), 631–649. doi: 10.1016/j.tplants.2021.03.010
- Wang, L., Zhou, X., Ren, X., Huang, L., Luo, H., Chen, Y., et al. (2018). A major and stable QTL for bacterial wilt resistance on chromosome B02 identified using a high-density SNP-based genetic linkage map in cultivated peanut yuanza 9102 derived population. *Front. Genet.* 9, 652. doi: 10.3389/fgene.2018.00652
- Zhao, C., He, L., Xia, H., Zhou, X., Geng, Y., Hou, L., et al. (2021). De novo full length transcriptome analysis of arachis glabrata provides insights into gene expression dynamics in response to biotic and abiotic stresses. *Genomics* 113 (3), 1579–1588. doi: 10.1016/j.ygeno.2021.03.030



Combining Genome-Wide Association Study and Gene-Based Haplotype Analysis to Identify Candidate Genes for Alkali Tolerance at the Germination Stage in Rice

Song Mei¹, Guogen Zhang², Jing Jiang¹, Jingbing Lu¹ and Fan Zhang^{1,2,3*}

¹Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China, ²College of Agronomy, Anhui Agricultural University, Hefei, China, ³National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya, China

OPEN ACCESS

Edited by:

Chuanzhi Zhao,
Shandong Academy of Agricultural
Sciences, China

Reviewed by:

Wricha Tyagi,
Central Agricultural University,
India
Jiang Shukun,
Heilongjiang Academy of Agricultural
Sciences, China

*Correspondence:

Fan Zhang
zhangfan03@caas.cn

Specialty section:

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

Received: 01 March 2022

Accepted: 22 March 2022

Published: 08 April 2022

Citation:

Mei S, Zhang G, Jiang J, Lu J and
Zhang F (2022) Combining Genome-
Wide Association Study and Gene-
Based Haplotype Analysis to Identify
Candidate Genes for Alkali Tolerance
at the Germination Stage in Rice.
Front. Plant Sci. 13:887239.
doi: 10.3389/fpls.2022.887239

Salinity–alkalinity stress is one of the main abiotic factors limiting rice production worldwide. With the widespread use of rice direct seeding technology, it has become increasingly important to improve the tolerance to salinity–alkalinity of rice varieties at the germination stage. Although we have a more comprehensive understanding of salt tolerance in rice, the genetic basis of alkali tolerance in rice is still poorly understood. In this study, we measured seven germination-related traits under alkali stress and control conditions using 428 diverse rice accessions. The alkali tolerance levels of rice germplasms varied considerably during germination. *Xian/indica* accessions had generally higher tolerance to alkali stress than *Geng/japonica* accessions at the germination stage. Using genome-wide association analysis, 90 loci were identified as significantly associated with alkali tolerance. Eight genes (*LOC_Os01g12000*, *LOC_Os03g60240*, *LOC_Os03g08960*, *LOC_Os04g41410*, *LOC_Os09g25060*, *LOC_Os11g35350*, *LOC_Os12g09350*, and *LOC_Os12g13300*) were selected as important candidate genes for alkali tolerance based on the gene functional annotation and gene-CDS-haplotype analysis. According to the expression levels of *LOC_Os09g25060* (*OsWRKY76*), it is likely to play a negative regulatory role in alkali tolerance during rice germination. An effective strategy for improving rice alkali tolerance may be to pyramid alkali-tolerant haplotypes of multiple candidate genes to obtain the optimal haplotype combination. Our findings may provide valuable genetic information and expand the use of alkali tolerance germplasm resources in rice molecular breeding to improve the alkali tolerance at the germination stage.

Keywords: alkali tolerance, germination stage, genome-wide association study, haplotype analysis, rice

INTRODUCTION

Salinity–alkalinity stress is considered to be one of the most severe abiotic stresses affecting crop growth and development, resulting in crop yield reduction (Qadir et al., 2014). Rice is one of the most important food crops in the world, and nearly a third of the world's population takes it as a staple food (Ray et al., 2013). Since the current rice yield level has reached a plateau,

improving the salinity–alkalinity tolerance of rice varieties and expanding the rice planting area is an effective strategy to increase the total yield. Usually, the increase of soluble salt in the soil is referred to as “soil salinization,” and soil salinization and alkalization are always simultaneous processes (Vengosh, 2014). Recent studies have found that their effects on plants are not exactly the same, and there are differences in how plants respond to their stress (Shi and Yin, 1993). Therefore, salt stress and alkali stress are two different stresses and should be treated separately. Salt stress is mainly caused by neutral salts, including NaCl and Na₂SO₄, while alkali stress is primarily caused by HCO₃[−] and CO₃^{2−} (Fang et al., 2021). Generally, salt stress causes ionic stress, osmotic stress, and oxidative stress to plants (Munns and Tester, 2008), while alkali stress is primarily caused by high pH (Wang et al., 2015) and may prove to be even more harmful to plants than salt stress (Yang et al., 2008). Rice is considered a salt- and alkali-sensitive crop (Cheng et al., 2007). The primary effects of saline–alkaline stress on rice are the decrease of germination rate, the inhibition or even death of seedling growth, and the decline of yield (Zhang et al., 2017). Nowadays, with rice direct seeding technology becoming more widely used, improving the salinity–alkalinity tolerance of rice varieties at the germination stage has become an important breeding target.

Salinity–alkalinity tolerance in rice is a complex trait controlled by multiple quantitative trait loci (QTLs; Liang et al., 2015). In recent years, many QTLs for salt tolerance in rice have been identified, and some salt tolerance genes, such as *SKC1* (Ren et al., 2005), *DST* (Huang et al., 2009), *OsCLC-1* (Nakamura et al., 2006), and *Saltol* (Thomson et al., 2010), have been cloned. In contrast, only a few QTLs are reported to influence alkali tolerance at the seedling stage in rice. Seven QTLs related to alkaline tolerance were detected using an F_{2,3} population derived from a cross between Caidao and WD20342 (Li et al., 2017). A rice alkali-tolerant mutant, called *alt1*, was identified and found that *alt1* negatively regulates alkali tolerance by preventing oxidative damage in rice (Guo et al., 2014). A calcium/calmodulin-dependent protein kinase *OsDMI3* can promote root elongation and improve alkaline tolerance of rice at germination stage by reducing Na⁺ and H⁺ influx in rice roots and maintaining ion balance (Ni et al., 2020).

Genome-wide association study (GWAS) has been widely used to identify associations between phenotypic traits and genotypes in many crops (Du et al., 2021; Ma et al., 2021a,b; Tekeu et al., 2021). Based on the GWAS using a core set of 208 rice germplasms, 6 and 14 SNPs associated with salt tolerance at the germination and seedling stages, and identified 22 candidate genes through haplotype analysis (Naveed et al., 2018). A major QTL (*RNC4*) related to root Na⁺/K⁺ ratio in *Geng/japonica* and *Xian/indica* accessions were identified by GWAS (Campbell et al., 2017). Yu et al. (2017) detected 25 SNPs associated with

salt tolerance at the seedling stage using 295 rice accessions by GWAS. Using 295 *Geng* accessions, eight QTLs associated with alkali tolerance at the seedling stage were detected by GWAS (Li et al., 2019). Li et al. jointly conducted QTL analysis and GWAS analysis on root length at the germination stage under alkali stress using 184 RILs and 295 *Geng* accessions and identified a major QTL *qAT11* (Li et al., 2020b). However, few studies have dissected the genetic basis of the alkali tolerance between different subspecies during rice germination.

In this study, a total of 428 accessions from the 3,000 rice genomes (3K-RG), mainly belonging to subspecies *Xian* and *Geng*, were evaluated for alkali tolerance at the germination stage (Wang et al., 2018). We used 2,949,726 SNPs filtered from the 4.8M SNP dataset in the Rice SNP-seek Database (Alexandrov et al., 2015) to identify QTLs and candidate genes associated with alkali tolerance by GWAS. This study aimed to understand the genetic basis of alkali tolerance during germination and provide favorable resources of genes and germplasms for molecular breeding for alkali tolerance improvement of rice.

MATERIALS AND METHODS

Plant Materials and Alkali Tolerance Evaluation

We used 428 accessions from the 3K-RG to evaluate the alkali tolerance of rice at the germination stage. The accessions contained 125 *Geng* accessions, 278 *Xian* accessions, 13 *Aus*, 6 *Basmati* and 6 *admix* accessions (Supplementary Table 1). A total of 120 seeds of each accession were dried at 50°C for 3 days to make them dry completely and break their dormancy. To disinfect the seeds, they were soaked in 3% NaClO solution for 30 min, and then the seeds were washed three times with sterile distilled water before testing. For each rice accession, 20 seeds were placed into a 90-mm-diameter Petri dish with filter paper. Three replicates were conducted under control (distilled water) and alkali stress conditions (0.15% Na₂CO₃ solution), respectively. As a control, 10 ml of distilled water was added to each Petri dish, and 10 ml of 0.15% Na₂CO₃ solution was added to each Petri dish as alkali stress. After that, the seeds were grown in a growth chamber under a 14-h light/10-h dark photoperiod (28°C/26°C) with 70% relative humidity for 7 days. Throughout the test, the solution was changed regularly every day.

Seeds were considered germinated when the root length was equal to the seed length and the shoot length was equal to half of the seed length. The germinated seeds were counted each day to calculate the germination rate (GR), germination energy (GE), and germination index (GI). We used the seed germination rates on day 3 and day 7 as GE and GR (GE or GR = N_t/N₀, where N_t represents the number of germinated seeds at day *t* and N₀ represents the total number of experimental seeds), respectively. The GI was calculated as follows: GI = $\sum (G_t/T_t)$, where G_t is the accumulated number of the germinated seeds on day *t* and T_t is the time corresponding to G_t in days (Wang et al., 2010). The mean germination time (MGT) was

Abbreviations: 3K-RG, 3,000 Rice genomes; MGT, Mean germination time; GR, Germination rate; GI, Germination index; VI, Vigor index; RL, Root length; SL, Shoot length; GE, Germination energy; ANOVA, Analysis of variance; GWAS, Genome-wide association study; MAF, Minor allele frequency; LD, Linkage disequilibrium; Chr, Chromosome; Kb, Kilobyte; Mb, Megabyte; SNP, Single nucleotide polymorphism; QTL, Quantitative trait locus/loci; Hap, Haplotype.

calculated by using the formula: $MGT = \sum T_i N_i / \sum N_i$, where N_i is the number of newly germinated seeds on day t (Alvarado et al., 1987). On day 7, the root length (RL) and shoot length (SL) of eight seeds randomly selected from each replicate were measured in both alkali stress and control conditions. Vigor index (VI) = (mean SL) \times GI. We also used the ratios of the seven germination-related traits under alkali stress to the control for evaluating the response of rice accessions to alkali stress. The average of three replicates for each trait was used for data analyses.

Genome-Wide Association Mapping

The 3K-RG 4.8M SNP dataset was downloaded from the Rice SNP-Seek Database (Alexandrov et al., 2015).¹ SNPs with the missing rate < 20% and minor allele frequencies > 5% were filtered with PLINK (Purcell et al., 2007). Finally, a total of 2,942,166, 2,091,233, and 1,218,609 SNPs were used for GWAS in the whole, *Xian* and *Geng* panel populations, respectively. The GWAS was performed with EMMAX (Kang et al., 2010) based on a mixed linear model (MLM) to detect the associations between SNP and the alkali tolerance-related traits. The kinship matrix was generated using an identical-by-state matrix (with the “emmax-kin -v -h -d 10” parameter in EMMAX) based on the pruned subset of genome-wide SNP data (with the “indep-pairwise 50 10 0.1” parameter in PLINK) to account for the relatedness among accessions. For controlling population structure, the eigenvectors of the kinship matrix were calculated using GCTA (with the “-make-grm” parameter; Yang et al., 2011) and the first three principal components were used as covariates. The effective number of independent markers (N) was calculated with the GEC software (Li et al., 2012) and suggestive significance thresholds of association by the Bonferroni correction method ($1/N$) were calculated for claiming significant SNPs for the whole population ($p = 2.15E-06$), *Xian* ($p = 2.76E-06$) and *Geng* ($p = 8.31E-06$) subpopulations, respectively. Manhattan and quantile–quantile (Q–Q) plots of GWAS were created by the R package “qqman” (Turner, 2014). Based on the previously reported linkage disequilibrium (LD) decay in the 3K-RG (Wang et al., 2018), the significant SNPs within a 300-kb region were considered a locus. A lead SNP in a locus was defined as the SNP with the lowest value of p , and the other significant SNPs within 150kb on either side of the lead SNP were merged.

Analysis of Candidate Genes

We shortlisted potential candidate genes for alkali tolerance when they met at least one of the following criteria: (1) genes with significant SNPs and functional annotation related to abiotic stress based on the Nipponbare reference genome IRGSP 1.0 (Kawahara et al., 2013) and the funRiceGenes database (Yao et al., 2018); (2) genes harboring SNPs significantly associated with more than three traits; (3) genes containing SNPs of the most significant association ($p < 0.05/\text{effective SNP number}$) with each trait. In the case of an SNP located in

the intergenic region, the downstream gene was chosen. LDBlockshow (Dong et al., 2021) was used to estimate the local LD block region containing each candidate gene. The haplotype analysis was performed on all candidate genes using all SNPs within the gene coding sequence region, in which synonymous SNPs were ignored (merged in one haplotype; Zhang et al., 2021). There were at least 10 rice accessions in each haplotype. Duncan’s multiple range post-hoc tests were completed with the “agricolae” package (de Mendiburu, 2021) in R to compare phenotypic differences among the haplotypes ($n \geq 10$ rice accessions). The expression profiles of candidate genes in Nipponbare were obtained from the RiceXPro database (Yutaka et al., 2013).

RNA Extraction and qRT-PCR Analysis

For detecting expression levels of candidate genes, we screened two representative accessions (one alkali-tolerant and one alkali-sensitive) from each subspecies. After 24h of alkali stress with 0.15% Na_2CO_3 , 100 seed embryos were sampled under alkali stress and control conditions, respectively. Total RNA was extracted from germinating seeds using plant RNA extraction kit (Tiangen Biotechnology) and reverse transcribed using reverse transcription kit (Invitrogen). Real-time qRT-PCR analyses were conducted with Taq Pro Universal SYBR qPCR Master Mix (Vazyme, Q712-02). All primers used for qRT-PCR are listed in **Supplementary Table 2**. The relative expression levels were calculated using the $2^{-\Delta\Delta\text{CT}}$ method (Livak and Schmittgen, 2002).

Statistical Analysis

We used one-way ANOVA to test the differences between subpopulations in the traits related to alkali tolerance by the “agricolae” (de Mendiburu, 2021) package in R. Spearman’s correlation coefficients between traits were calculated using the “corrplot” package in R. Box plots for displaying the trait distribution and bar charts for illustrating haplotype frequency were generated using the “ggplot2” (Ginestet, 2011) package in R.

RESULTS

Phenotypic Variations in the Alkali Tolerance of Rice at the Germination Stage

Seven traits including GR, GI, MGT, VI, RL, SL, and GE were measured for 428 rice accessions under alkali stress with 0.15% Na_2CO_3 (named “TraitName+S”) and control condition with distilled water (named “TraitName+C”) at the germination stage (**Table 1**). Furthermore, we calculated the ratios of these traits under alkali stress and control conditions (named “R+TraitName”) in order to determine the degree of alkali damage (**Supplementary Table 1**). Under alkali stress and control conditions, the germination-related traits were distributed continuously, suggesting that multiple loci may be involved in controlling these traits (**Supplementary Figure 1**). We observed 12 accessions with GRS of 1.00, including eight *Xian* accessions, three *Geng* accessions, and an *Aus* accession

¹<http://snp-seek.irri.org/>

TABLE 1 | Summary of the 21 traits measured in this study.

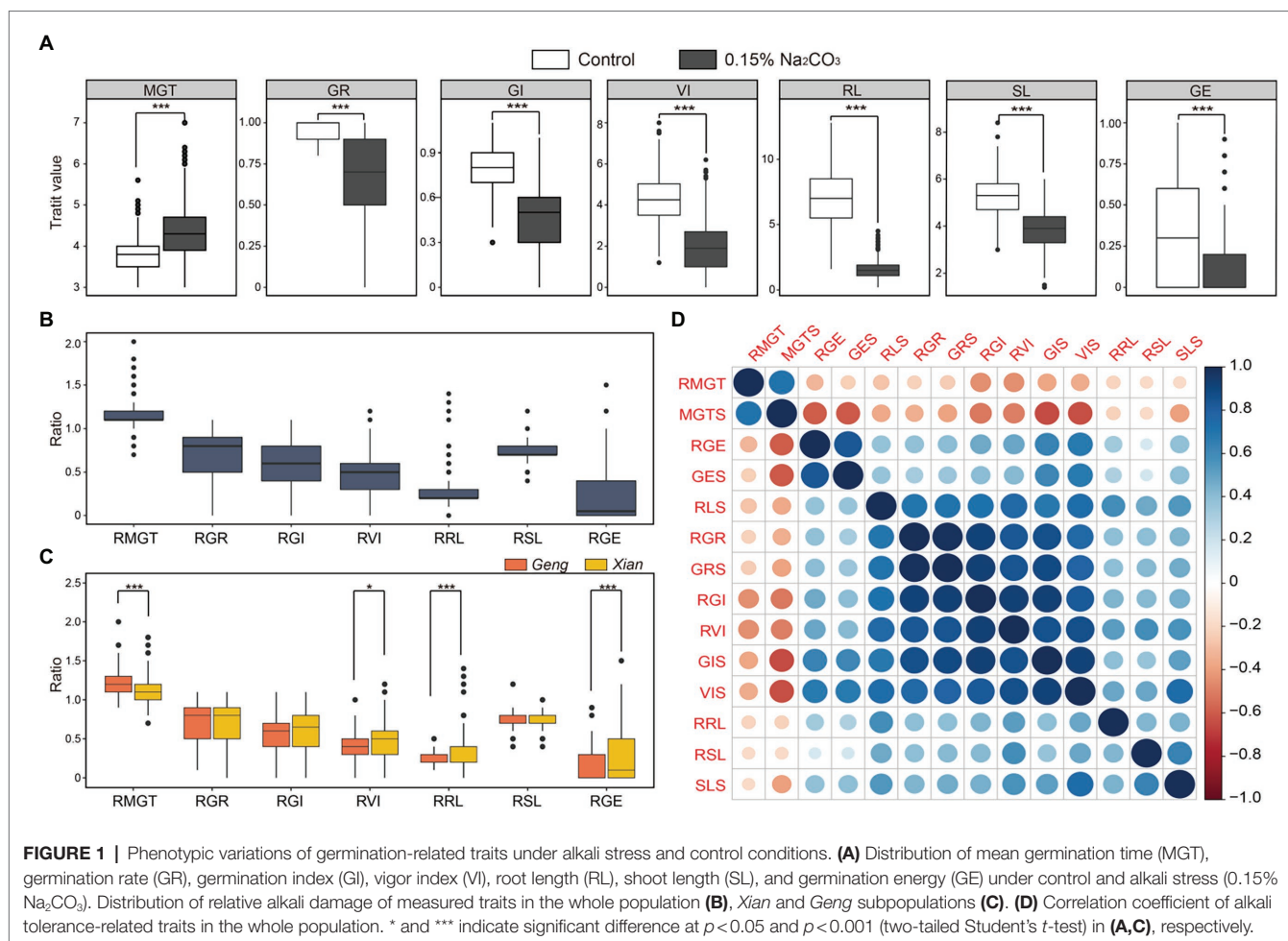
Trait	Full names of traits	Range	Mean	SD	CV
MGTS	Mean germination time under alkali stress (d)	3.000–7.000	4.337	0.698	0.161
MGTC	Mean germination time under control condition (d)	3.000–5.556	3.785	0.434	0.115
GRS	Germination rate under alkali stress	0.017–1.000	0.647	0.266	0.411
GRC	Germination rate under control condition	0.800–1.000	0.955	0.048	0.050
GIS	Germination index under alkali stress	0.007–1.040	0.471	0.244	0.518
GIC	Germination index under control condition	0.343–1.093	0.809	0.136	0.168
VIS	Vigor index under alkali stress	0.013–6.173	1.966	1.230	0.625
VIC	Vigor index under control condition	1.235–8.016	4.323	1.160	0.268
RLS	Root length under alkali stress (cm)	0.163–4.525	1.582	0.679	0.429
RLC	Root length under control condition (cm)	1.600–12.821	6.858	2.215	0.323
SLS	Shoot length under alkali stress (cm)	1.363–6.025	3.887	0.802	0.206
SLC	Shoot length under control condition (cm)	3.018–8.383	5.293	0.791	0.150
GES	Germination energy under alkali stress	0.000–0.900	0.128	0.207	1.619
GEC	Germination energy under control condition	0.000–1.000	0.329	0.316	0.963
RMGT	Ratio of mean germination time under alkali stress to normal condition	0.735–2.004	1.149	0.158	0.138
RGR	Ratio of germination rate under alkali stress to normal condition	0.017–1.118	0.678	0.277	0.409
RGI	Ratio of germination index under alkali stress to normal condition	0.008–1.127	0.574	0.266	0.463
RVI	Ratio of vigor index under alkali stress to normal condition	0.004–1.232	0.441	0.227	0.514
RRL	Ratio of root length under alkali stress to normal condition	0.013–1.406	0.264	0.170	0.645
RSL	Ratio of shoot length under alkali stress to normal condition	0.371–1.212	0.733	0.100	0.136
RGE	Ratio of germination energy under alkali stress to normal condition	0.000–1.500	0.211	0.283	1.338

Range, Mean, SD, and CV indicate the range from minimum value to maximum value, mean value, standard deviation, and coefficient of variation of a trait, respectively.

(**Supplementary Table 1**). Among the four accessions with the shortest MGTS, all were *Xian* accessions. In addition, we found some accessions exhibited extreme phenotypes in other germination-related traits, such as “Lumaozhan” with the highest GES (0.90) and GIS (1.04), “Heimangdao” with the longest SLS (6.03 cm), “ARC 15091” with the longest RLS (4.53 cm), and “NCS 766” with the highest VIS (6.17).

In comparing the germination-related traits under alkali stress and control conditions, we found that alkali stress significantly impaired the seed germination and growth, extending

germination time and reducing GE, GR, GI, VI, RL, and SL (**Figure 1A**). GE was the most strongly affected by alkali stress among the seven traits. On average, MGT was delayed by 14.9% under alkali stress, and GR, GI, VI, RL, SL, and GE decreased by 32.2%, 42.6%, 55.9%, 73.6%, 26.7%, and 78.9%, respectively. We used the ratio of the trait value under alkali stress to control to determine the extent of damage caused by alkali. Accordingly, alkali damage on these germination-related traits followed the following order: GE > RL > VI > GI > GR > SL > MGT (**Figure 1B**). Furthermore,



the *Xian* subpopulation ($n=278$) exhibited significantly higher RVI, RRL, RGE, but significantly shorter RMGT than the *Geng* subpopulation ($n=125$; **Figure 1C**), suggesting that *Xian* accessions are generally more alkaline tolerant than *Geng* accessions.

According to the correlation analysis among the germination-related traits in the whole population, MGTS and RMGT were negatively correlated with the other traits, whereas GRS, RGR, GIS, RGI, VIS, and RVI demonstrated a strong positive correlation with one another (**Figure 1D**). In addition, correlation analysis results in the subpopulations *Xian* and *Geng* were similar to the results in the whole population (**Supplementary Figure 2**). These findings indicate that it is valuable to identify the loci associated with these germination traits and corresponding favorable alleles in molecular breeding for rice alkali tolerance.

Identification of Loci Associated With Alkali Tolerance

We conducted a GWAS based on the mixed linear model for the traits related to alkali tolerance (**Supplementary Figures 3–8**). Using a Bonferroni correction based on the effective numbers, the genome-wide significant value of p thresholds were set at $2.15\text{E}-06$, $2.76\text{E}-06$, and $8.31\text{E}-06$ for the whole population,

Xian subpopulation, and *Geng* subpopulation, respectively. Consequently, 271, 51, and 503 SNPs significantly associated with alkali tolerance at the germination stage were identified in the whole population, *Xian* subpopulation and *Geng* subpopulation, respectively. A total of 18 significantly associated SNPs were detected both in the whole population and *Xian* subpopulation (**Supplementary Figure 9A**). However, there were no overlaps between the identified SNPs in *Xian* subpopulation and those in *Geng* subpopulation. By analyzing the genes in which the significantly associated SNPs were detected, it was found that there was also no overlap between the genes identified in *Xian* subpopulation and those in *Geng* subpopulation (**Supplementary Figure 9B**). The results suggested that the genetic mechanism of alkali tolerance in subpopulations *Xian* and *Geng* may be different. We merged the adjacent SNPs around 300 kb as a locus to reduce the redundancy in association signals of different traits. Consequently, a total of 90 loci involved in 124 associations between 106 lead SNPs and 14 traits were identified in the whole population or at least one of the two subpopulations (**Supplementary Table 3**).

According to the annotations of the genes with identified SNPs, no reported genes for alkaline tolerance were identified. However, we found two known salt tolerance genes (*OsRLCK253* and *OsNCED5*) harboring lead SNPs (rs8_17550068 with

$p=4.92\text{E-}07$ and rs12_26284167 with $p=6.1\text{E-}06$ associated with RSL in *Geng* subpopulation (Figure 2A). Moreover, a WRKY transcription factor gene *OsWRKY76* was significantly associated with RRL in the whole population (Figure 2B). In addition, we identified several new loci with strong associations with alkali tolerance (Figures 2C–E; Supplementary Table 3). For example, a lead SNP rs5_26741198 was significantly associated with SLS ($p=8.68\text{E-}11$) in the whole population. The lead SNP of rs3_34261591 showed the strongest association signal ($p=2.15\text{E-}8$) with MGTS in *Xian* subpopulation. The lead SNP rs8_24108555 with the most significant association ($p=9.42\text{E-}9$) for GES was identified in *Geng* subpopulation (Figures 2C–E).

A comprehensive analysis was carried out concerning the 90 alkali tolerance loci, each of which was associated with one to five traits (Figure 2F; Supplementary Table 3). For example, locus 4 was associated with GRS, RLS, RGR, and RGI in the whole population and with RLS in *Geng* subpopulation. Locus 7 was associated with GRS, RGR, and RVI in the whole

population. Locus 44 was associated with GES, RGE and RRL in the whole population and with RRL in *Xian* subpopulation. Locus 59 was associated with GRS and RGR both in the whole population and *Xian* subpopulation. Locus 88 was associated with GIS, VIS, SLS, and RVI in the whole population and with RSL in *Geng* subpopulation.

Haplotype Analyses of the Candidate Genes

We selected 17 genes as candidate genes at 15 loci (Supplementary Table 4; see Methods for detail), and conducted a gene-CDS-haplotype analysis to determine whether there were significant differences in alkali tolerance between different haplotypes of each gene. Finally, eight genes with significant phenotypic differences between haplotypes were identified as important candidate genes for alkali tolerance, including *LOC_Os09g25060*, *LOC_Os03g60240*, *LOC_Os03g08960*, *LOC_Os01g12000*, *LOC_Os04g41410*, *LOC_Os11g35350*,

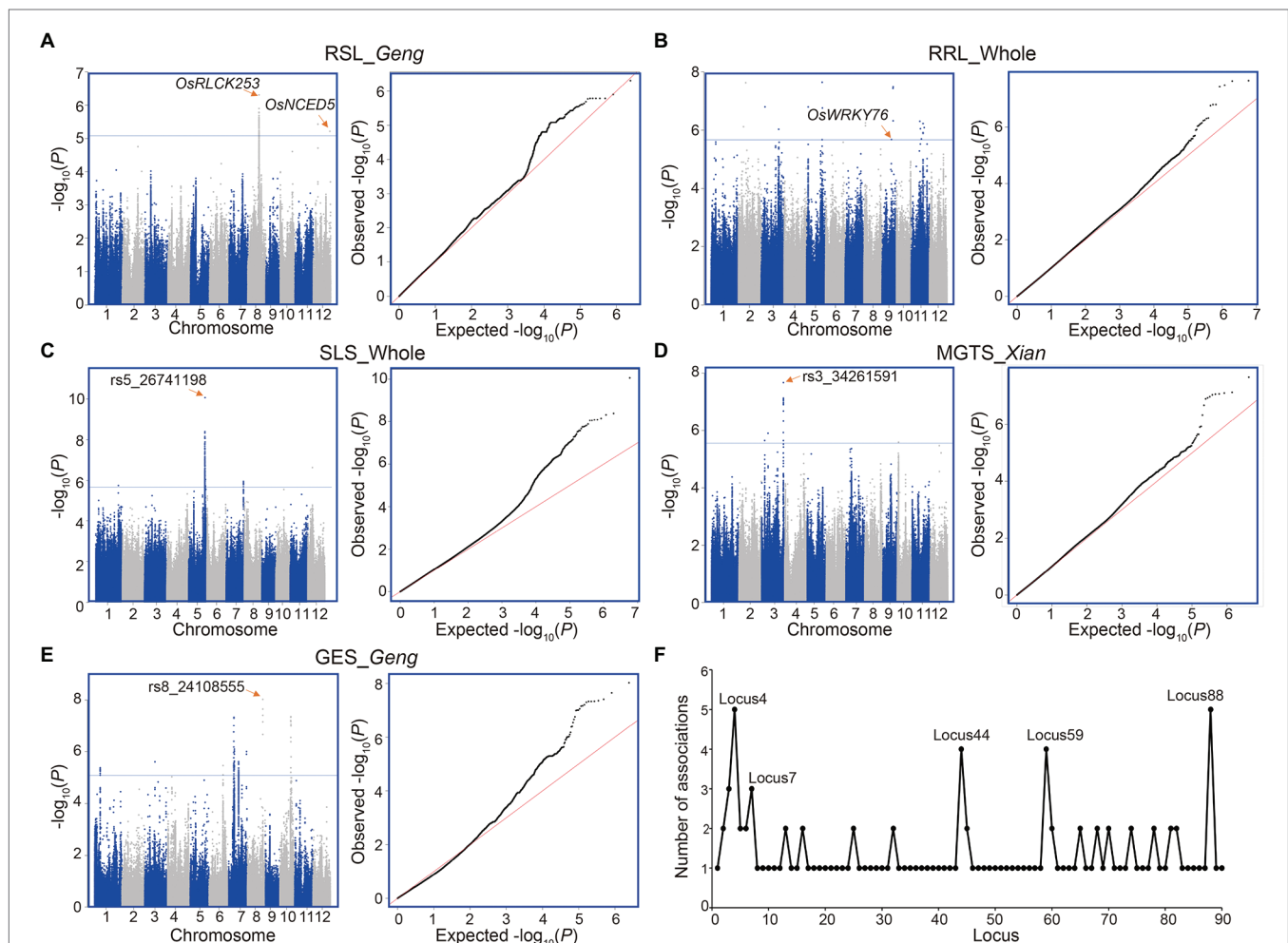


FIGURE 2 | Genome-wide association study of five alkali tolerance-related traits. (A) RSL in the *Geng* subpopulation. (B) RRL in the whole population. (C) SLS in the whole population. (D) MGTS in the *Xian* subpopulation. (E) GES in the *Geng* subpopulation. (F) Number of associated traits for each of all detected loci (Supplementary Table 3). Manhattan plot (left) and Q-Q plot (right) for each panel in (A–E). Horizontal lines indicate in the Manhattan plots indicate the genome-wide suggestive thresholds. The names of candidate genes or lead SNPs are shown above the corresponding association signals.

LOC_Os12g09350, and *LOC_Os12g13300*. Based on qRT-PCR analysis, the expression levels of *LOC_Os09g25060* and *LOC_Os03g60240* were both higher in the alkali-sensitive *Xian* accession than in the alkali-tolerant *Xian* accession. On the contrary, *LOC_Os11g35350* was expressed more strongly in the alkali-tolerant *Xian* accession compared to the alkali-sensitive *Xian* accession. There were no significant differences in the expression levels of the other five genes between the two *Xian* accessions (Figure 3A). For the two *Geng* accessions, the expression level of *LOC_Os09g25060* in alkali-sensitive accession was also significantly higher than that in alkali-tolerant accession, while *LOC_Os01g12000* was expressed more strongly in the alkali-tolerant accession than in alkali-sensitive accession (Figure 3B).

For locus 71 on chromosome 9, the LD block region was predicted from 14.94 to 14.99 Mb (48.93 kb) and included 491 SNPs (Figure 4A). *LOC_Os09g25060* (*OsWRKY76*) had a lead SNP rs9_14985169 ($p = 2.09E-06$) for RRL in the whole subpopulation, which is a member of the rice WRKY transcription factor gene family (Ross et al., 2007). Three major haplotypes of *LOC_Os09g25060* were detected based on five SNPs in the coding region shared by at least 10 accessions (Figure 4B). Hap1 is predominant in *Xian* subpopulation and Hap2 is prevalent in *Geng* subpopulation (Figure 4D). In comparison of the RRL across the three haplotypes, Hap1 and Hap2 had a significantly higher RRLs of 0.28 and 0.26 than Hap3, which had a mean RRL of 0.19 in the whole population (Figure 4C).

The LD block region for locus 32 on chromosome 3 was predicted from 34.256 to 34.267 Mb (11.16 kb) and contained 37 SNPs (Figure 5A), including the alkali tolerance candidate gene *LOC_Os03g60240* that had the most significant SNP rs3_34257992 ($p = 7.96E-08$) for MGTS in *Xian* subpopulation. Using 22 SNPs (including five non-synonymous SNPs) in the coding region of *LOC_Os03g60240*, six haplotypes were detected, each of which was shared by at least 10 accessions (Figure 5B).

In the whole population, Hap4 had the lowest mean MGTS (3.98). Thus, we defined Hap4 as the favorable haplotypes of alkali tolerance, which were carried only by *Xian* accessions (Figures 5C,D).

For locus 7 on chromosome 1, the most significant association peak for RGR, GRS, and RVI in the whole population, was predicted from 6.47 to 6.56 Mb (85.5 kb) and included 730 SNPs based on local LD analysis (Figure 6A). One of these SNPs, rs1_6533274, located in the intergenic region of two genes (*LOC_Os01g11990* and *LOC_Os01g12000*), was significantly associated with RGR, GRS, and RVI. We identified four haplotypes ($n \geq 10$ accessions) based on eight SNPs in the coding region of *LOC_Os01g12000*, the downstream gene of rs1_6533274 (Figure 6B). Among the haplotypes, Hap1 and Hap3 were both enriched in *Xian* subpopulation, while Hap3 had significantly higher GRS, RGR, and RVI than the other haplotypes (Figures 6C,D).

The LD block region of locus 82 was estimated to be from 20.71 to 20.75 (38.1 kb) on chromosome 11 including 438 SNPs. The SNP rs11_20724735, located in the candidate gene *LOC_Os11g35350*, was significantly associated with RRL in the *Xian* subpopulation (Figure 7A). Using six SNPs within the coding region of *LOC_Os11g35350*, four major haplotypes ($n \geq 10$ accessions) were identified (Figure 7B). Hap3 had significantly higher RRL, which were mainly present in the *Xian* accessions compared to the other haplotypes (Figures 7C,D).

Optimal Alkali-Tolerant Haplotype Combination

Since RGE was most sensitive to alkali stress, it was used to identify the favorable haplotype of each candidate gene. As a result, Hap1 of *LOC_Os09g25060*, Hap3 of *LOC_Os01g12000*, Hap5 of *LOC_Os12g13300*, Hap4 of

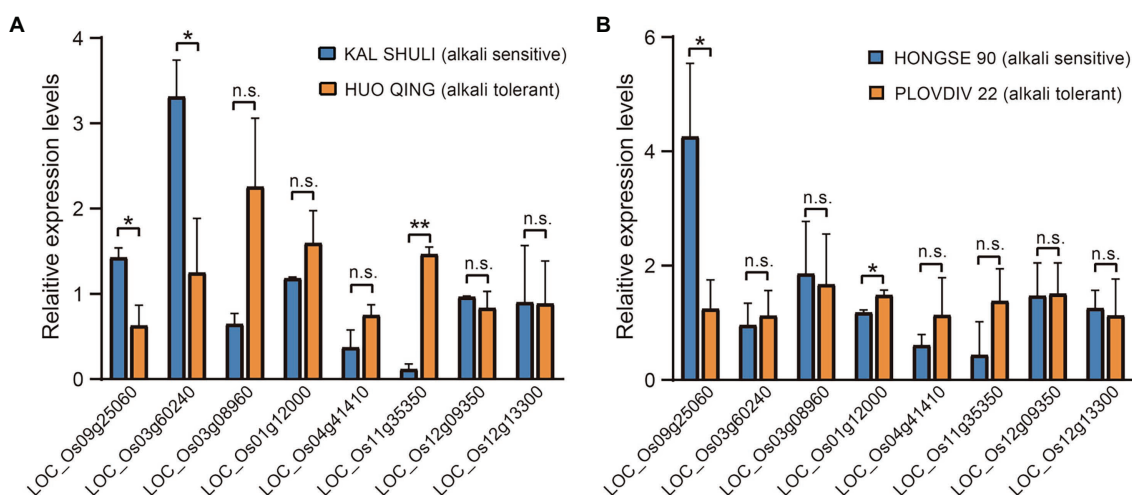


FIGURE 3 | Relative expression levels of candidate genes under alkali stress for 24 h. (A) In two representative *Xian* germplasms. (B) In two representative *Geng* accessions. Gene expression was normalized to that of the *OsActin* gene control. The relative expression levels were represented by fold change relative to the expression levels of the candidate genes. * $p < 0.05$ and ** $p < 0.01$ (two-tailed Student's *t*-test). Data represent means \pm SD ($n = 3$).

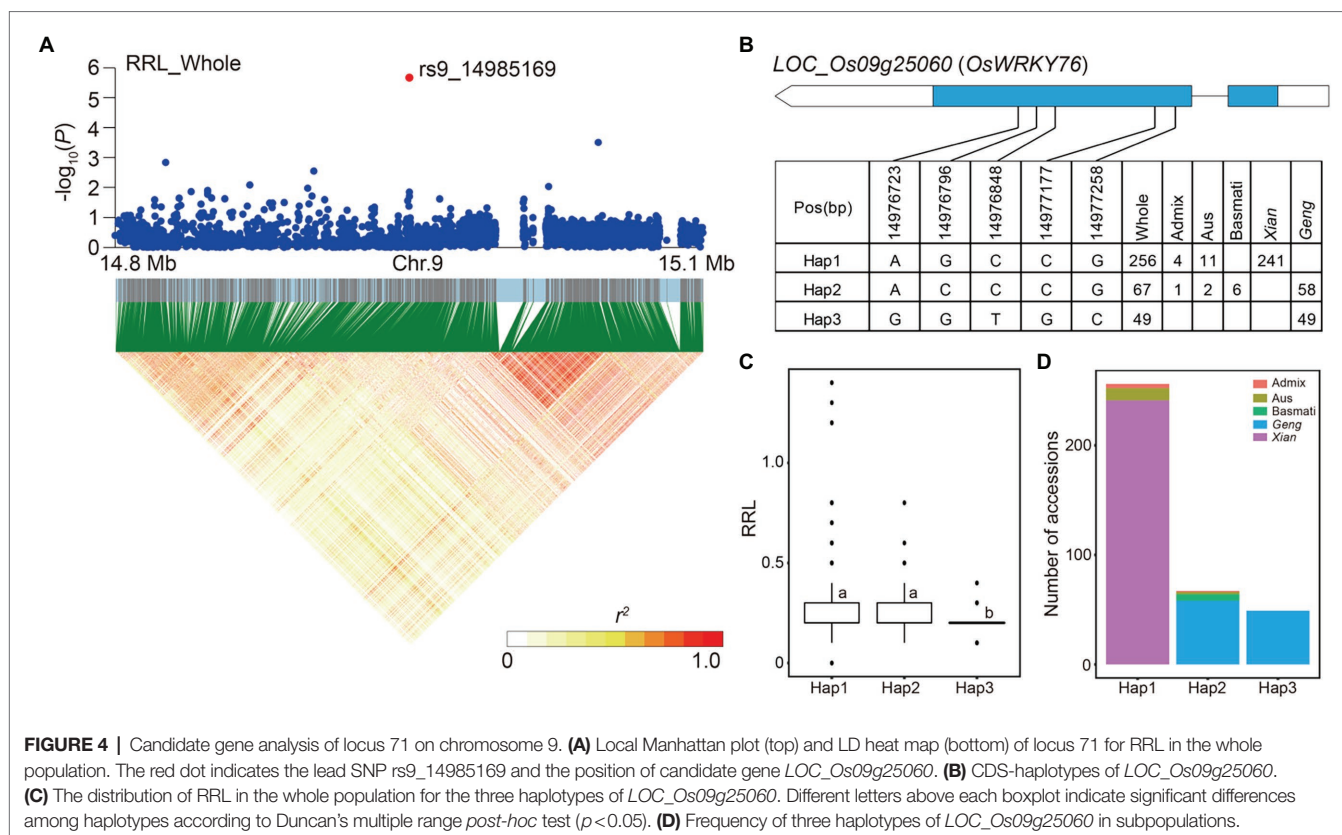


FIGURE 4 | Candidate gene analysis of locus 71 on chromosome 9. **(A)** Local Manhattan plot (top) and LD heat map (bottom) of locus 71 for RRL in the whole population. The red dot indicates the lead SNP rs9_14985169 and the position of candidate gene *LOC_Os09g25060*. **(B)** CDS-haplotypes of *LOC_Os09g25060*. **(C)** The distribution of RRL in the whole population for the three haplotypes of *LOC_Os09g25060*. Different letters above each boxplot indicate significant differences among haplotypes according to Duncan's multiple range *post-hoc* test ($p < 0.05$). **(D)** Frequency of three haplotypes of *LOC_Os09g25060* in subpopulations.

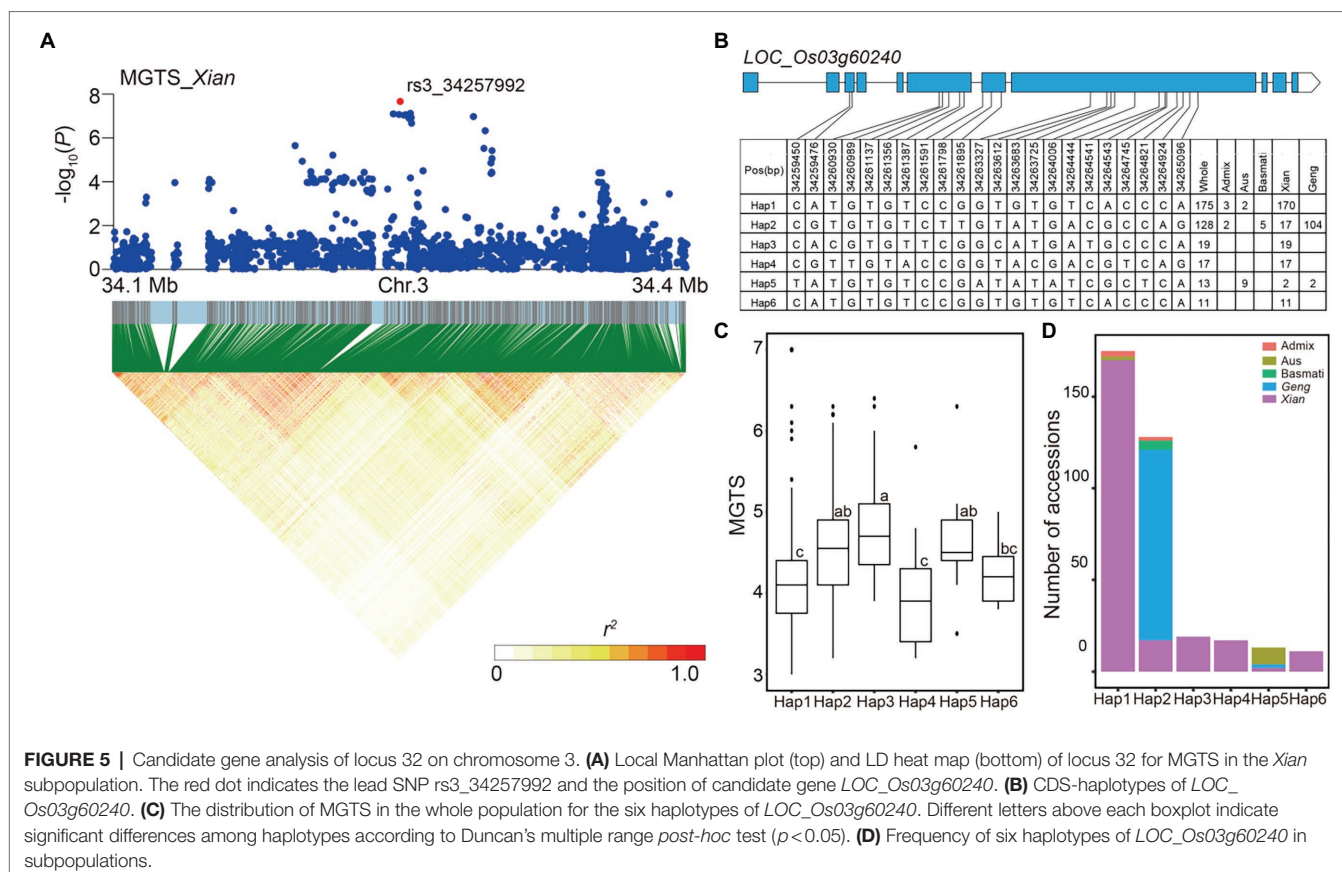


FIGURE 5 | Candidate gene analysis of locus 32 on chromosome 3. **(A)** Local Manhattan plot (top) and LD heat map (bottom) of locus 32 for MGTS in the *Xian* subpopulation. The red dot indicates the lead SNP rs3_34257992 and the position of candidate gene *LOC_Os03g60240*. **(B)** CDS-haplotypes of *LOC_Os03g60240*. **(C)** The distribution of MGTS in the whole population for the six haplotypes of *LOC_Os03g60240*. Different letters above each boxplot indicate significant differences among haplotypes according to Duncan's multiple range *post-hoc* test ($p < 0.05$). **(D)** Frequency of six haplotypes of *LOC_Os03g60240* in subpopulations.

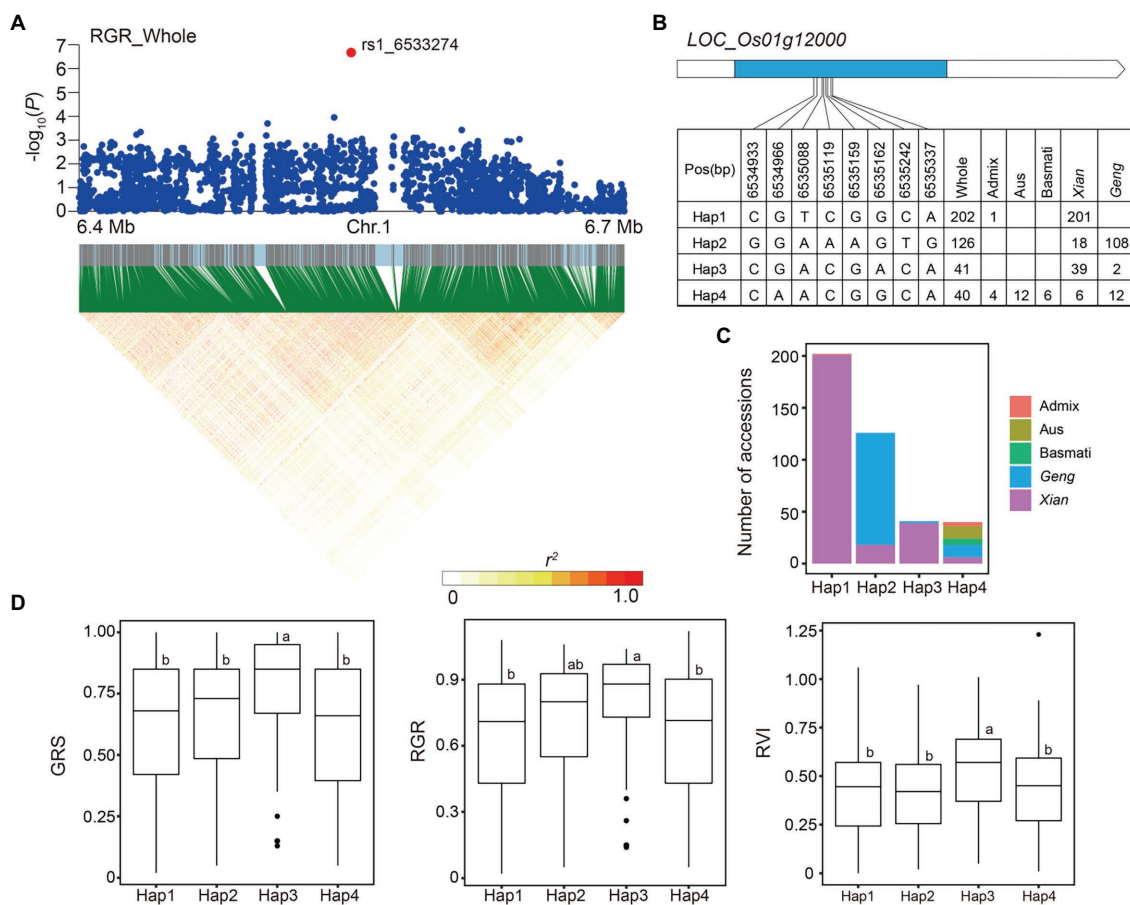


FIGURE 6 | Candidate gene analysis of locus 7 on chromosome 1. **(A)** Local Manhattan plot (top) and LD heat map (bottom) of locus 7 for RGR in the whole population. The red dot indicates the lead SNP rs1_6533274 and the position of candidate gene *LOC_Os01g12000*. **(B)** CDS-haplotypes of *LOC_Os01g12000*. **(C)** Frequency of four haplotypes of *LOC_Os01g12000* in subpopulations. **(D)** The distribution of GRS, RGR, and RVI in the whole population for the four haplotypes of *LOC_Os01g12000*. Different letters above each boxplot indicate significant differences among haplotypes according to Duncan's multiple range *post-hoc* test ($p < 0.05$).

LOC_Os03g60240, Hap3 of *LOC_Os04g41410*, Hap6 of *LOC_Os12g09350*, Hap4 of *LOC_Os11g35350*, and Hap1 of *LOC_Os03g08960* were detected as alkali-tolerant haplotypes. We used a combined haplotype analysis in order to determine the breeding potential of candidate genes for alkali tolerance. There remained five groups comprising three candidate genes after the removal of rare haplotype combinations ($n < 15$ accessions). The Groups I, II, III, and IV were enriched in the *Xian* subpopulation, whereas Group V was mainly found in the *Geng* subpopulation, which showed clear *Xian-Geng* differentiation (Figure 8A). Compared to the other four groups, Group I containing the alkali-tolerant haplotypes at *LOC_Os03g60240*, *LOC_Os09g25060*, and *LOC_Os03g08960* had a significantly higher RGE (Figure 8B). Furthermore, the more favorable haplotypes of these candidate genes were pyramided in rice, the stronger the alkali tolerance. The results suggest that developing a pyramid of multiple favorable haplotypes/alleles for alkali tolerance may be an effective strategy for improving rice alkali tolerance at the germination stage.

DISCUSSION

Xian Is More Alkali-Tolerant Than *Geng* During Germination

The alkali stress of 0.15% Na_2CO_3 can significantly inhibit rice germination (Figure 1A). Qi et al. (Qi et al., 2005, 2006) proposed the primary method to evaluate the alkali tolerance during rice germination is to measure GR, GI, and relative alkali damage rate, and they reported a wide variation range and reasonable distribution of GR, GE and their relative alkali damage rate under 0.10% and 0.20% Na_2CO_3 condition. In our study, GE was the most sensitive trait of the seven germination-related traits measured under alkali stress (Figure 1B), which is consistent with previous studies (Lv et al., 2013). Furthermore, the alkali tolerance levels of rice germplasms also varied considerably during germination in this study (Supplementary Figure 1). *Xian* accessions had shorter RMGT and higher RVI, RRL, and RGE than *Geng* accessions, indicating that *Xian* accessions had generally higher tolerance to alkali stress than *Geng* accessions at the germination

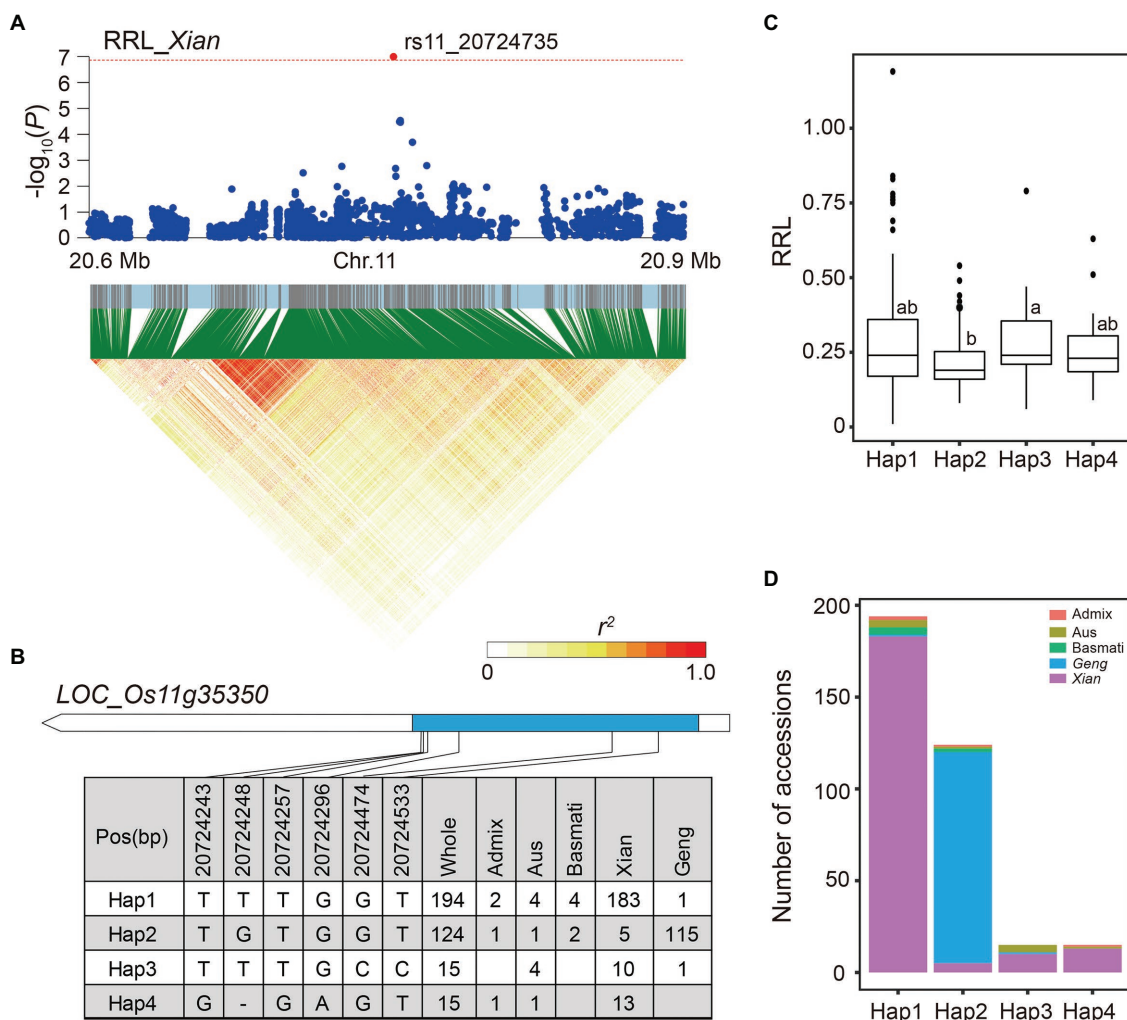


FIGURE 7 | Candidate gene analysis of locus 82 on chromosome 11. **(A)** Local Manhattan plot (top) and LD heat map (bottom) of locus 82 for RRL in the *Xian* subpopulation. The red dot indicates the lead SNP rs11_20724735 and the position of candidate gene *LOC_Os11g35350*. **(B)** CDS-haplotypes of *LOC_Os11g35350*. **(C)** The distribution of RRL in the whole population for the four haplotypes of *LOC_Os11g35350*. Different letters above each boxplot indicate significant differences among haplotypes according to Duncan's multiple range *post-hoc* test ($p < 0.05$). **(D)** Frequency of four haplotypes of *LOC_Os11g35350* in subpopulations.

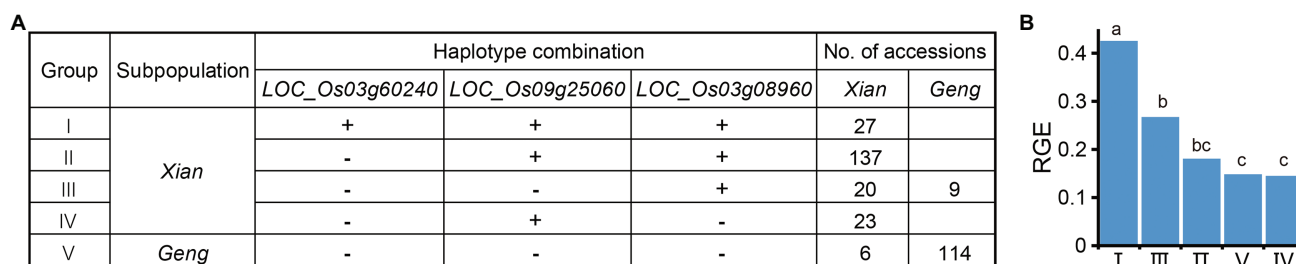


FIGURE 8 | Optimal alkali-tolerant haplotype combination of candidate genes. **(A)** Combined haplotypes of *LOC_Os03g60240*, *LOC_Os09g25060*, and *LOC_Os03g08960*. "+" and "-" represent favorable and inferior haplotypes, respectively. **(B)** Comparison of the RGE among different haplotype combinations. Different letters above each histogram indicate significant differences at $p < 0.05$ (Least Significant Difference test).

stage (Figure 1C). This is contrary to the previous findings by Shi et al. regarding salt tolerance in rice during germination

(Shi et al., 2017), inferring different mechanisms for salt and alkali tolerance in rice.

Genetic Mechanism Underlying Rice Alkali Tolerance Differs From Salt Tolerance During Germination

Two salt tolerance-related genes *OsRLCK253* and *OsNCED5* were associated with alkali tolerance at the germination stage in this study. *OsRLCK253*, which encodes a receptor-like cytoplasmic kinase, can improve the water-deficit and salt tolerance in transgenic *Arabidopsis* plants by affecting several common endogenous gene expressions (Giri et al., 2011). *OsNCED5*, encoding a 9-cis-epoxycarotenoid dioxygenase, positively regulates ABA level, enhances salt tolerance, and accelerates leaf senescence in rice (Xiong et al., 2014). In both *OsRLCK253* and *OsNCED5*, however, there were no significant differences in alkali tolerance between different haplotypes. In addition, no cloned genes with large effects on salt tolerance were detected in this study, further supporting the perspective that different genetic mechanisms may regulate alkali tolerance and salt tolerance in rice (Sun et al., 2021).

The expression levels of candidate gene *LOC_Os09g25060* were significantly higher in alkali-sensitive accessions compared with alkali-tolerant accessions in both *Xian* and *Geng* subpopulations, suggesting its negative regulatory role in alkali tolerance during rice germination. *LOC_Os09g25060*, encoding a WRKY transcription factor *OsWRKY76*, is downregulated in the roots of Dongdao-4 in response to saline-alkaline stress (Li et al., 2020a). Previous studies have shown that the genes associated with jasmonic acid (JA) biosynthesis and JA signaling are enhanced in the knockout plants of both *OsWRKY62* and *OsWRKY76* (Liang et al., 2017). Moreover, salt stress can trigger the activation of the JA signaling pathway and inhibit the cell elongation in the elongation zone of root (Valenzuela et al., 2016). Based on haplotype analysis of *OsWRKY76*, Hap1 and Hap2 had higher RRL than Hap3 (Figure 4C), suggesting that *OsWRKY76* may affect root development by activating the JA signaling pathway, which ultimately affects rice alkali tolerance at the germination stage.

Breeding Application of the Favorable Haplotypes

Rice can alleviate the damage of alkali stress through osmotic regulation, ion balance, antioxidant protection, and hormone regulation (Fang et al., 2021). The alkali tolerance of rice at the germination stage is a complex trait, which is controlled by multiple genes.

Among 90 loci for alkali tolerance in this study, only five loci (5.6%) were also detected in previous QTL mapping studies (Supplementary Table 3), primarily because most reported QTLs for alkali tolerance were identified at the seedling stage. In other words, the genetic mechanisms of alkali tolerance in

rice may partially differ between the seedling and germination stages. Our results indicated that rice germplasms with more favorable haplotypes on candidate genes had better alkali tolerance (Figure 8), which suggests that developing an optimal haplotype combination through pyramiding multiple favorable haplotypes should be an effective strategy for improving alkali tolerance during rice germination. In addition, it is possible, through inter-subspecific crosses in rice breeding programs, to make introgression of favorable haplotypes specific in the *Xian* subpopulation into the *Geng* subpopulation to provide important donor resources for improving the alkali tolerance of the *Geng* subpopulation. For example, rice breeders may use marker-assisted selection for Hap4 of *LOC_Os03g60240* in the progeny of inter-subspecific crosses to shorten germination time under alkali stress. Thus, the candidate genes associated with alkali tolerance and their favorable haplotypes will be useful to rice alkali-tolerant breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, and further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

FZ conceived and designed the experiments. SM analyzed the data and wrote the manuscript. SM, GZ, JJ, and JL performed the experiments. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key Research and Development Project of Hainan Province (ZDYF2021XDNY171), and the Agricultural Science and Technology Innovation Program and the Cooperation and Innovation Mission (CAAS-ZDXT 202001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.887239/full#supplementary-material>

REFERENCES

- Alexandrov, N., Tai, S. S., Wang, W. S., Mansueto, L., Palis, K., Fuentes, R. R., et al. (2015). SNP-seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* 43, D1023–D1027. doi: 10.1093/nar/gku1039
- Alvarado, A. D., Bradford, K. J., and Hewitt, J. D. (1987). Osmotic priming of tomato seed: effect on germination, field emergence, seedling growth and fruit yield. *J. Am. Soc. Hortic. Sci.* 112, 427–432.
- Campbell, M. T., Bandillo, N., Al Shiblawi, F. R. A., Sharma, S., Liu, K., Du, Q., et al. (2017). Allelic variants of *OsHKT1;1* underlie the divergence between indica and japonica subspecies of rice (*Oryza sativa*) for root sodium content. *PLoS Genet.* 13:e1006823. doi: 10.1371/journal.pgen.1006823
- Cheng, S. H., Cao, L. Y., Zhuang, J. Y., Chen, S. G., Zhan, X. D., Fan, Y. Y., et al. (2007). Super hybrid rice breeding in China: achievements and prospects. *J. Integr. Plant Biol.* 49, 805–810. doi: 10.1111/j.1744-7909.2007.00514.x

- de Mendiburu, F. (2021). *Agricolae*: statistical procedures for agricultural research. R Package Version 1.3-5. Available at: <https://cran.r-project.org/web/packages/agricolae> (Accessed January 12 2022).
- Dong, S. S., He, W. M., Ji, J. J., Zhang, C., Guo, Y., and Yang, T. L. (2021). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* 22:bbaa227. doi: 10.1093/bib/bbaa227
- Du, X., Xu, W., Peng, C., Li, C., Zhang, Y., and Hu, L. (2021). Identification and validation of a novel locus, Qpm-3BL, for adult plant resistance to powdery mildew in wheat using multilocus GWAS. *BMC Plant Biol.* 21:357. doi: 10.1186/s12870-021-03093-4
- Fang, S., Hou, X., and Liang, X. (2021). Response mechanisms of plants Under saline-alkali stress. *Front. Plant Sci.* 12:667458. doi: 10.3389/fpls.2021.667458
- Ginestet, C. (2011). ggplot2: elegant graphics for data analysis. *J. R. Stat. Soc. A. Stat. Soc.* 174, 245–246. doi: 10.1111/j.1467-985X.2010.00676_9.x
- Giri, J., Vij, S., Dansana, P. K., and Tyagi, A. K. (2011). Rice A20/AN1 zinc-finger containing stress-associated proteins (SAP1/11) and a receptor-like cytoplasmic kinase (OsRLCK253) interact via A20 zinc-finger and confer abiotic stress tolerance in transgenic Arabidopsis plants. *New Phytol.* 191, 721–732. doi: 10.1111/j.1469-8137.2011.03740.x
- Guo, M., Wang, R., Wang, J., Hua, K., Wang, Y., Liu, X., et al. (2014). ALT1, a Snf2 family chromatin remodeling ATPase, negatively regulates alkaline tolerance through enhanced defense against oxidative stress in Rice. *PLoS One* 9:e112515. doi: 10.1371/journal.pone.0112515
- Huang, X. Y., Chao, D. Y., Gao, J. P., Zhu, M. Z., Shi, M., and Lin, H. X. (2009). A previously unknown zinc finger protein, DST, regulates drought and salt tolerance in rice via stomatal aperture control. *Genes Dev.* 23, 1805–1817. doi: 10.1101/gad.1812409
- 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
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. doi: 10.1186/1939-8433-6-4
- Li, M. X., Yeung, J. M., Cherny, S. S., and Sham, P. C. (2012). Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* 131, 747–756. doi: 10.1007/s00439-011-1118-2
- Li, N., Sun, J., Wang, J. G., Liu, H. L., Zheng, H. L., Yang, L. M., et al. (2017). QTL analysis for alkaline tolerance of rice and verification of a major QTL. *Plant Breed.* 136, 881–891. doi: 10.1111/pbr.12539
- Li, N., Zheng, H., Cui, J., Wang, J., Liu, H., Sun, J., et al. (2019). Genome-wide association study and candidate gene analysis of alkalinity tolerance in japonica rice germplasm at the seedling stage. *Rice* 12:24. doi: 10.1186/s12284-019-0285-y
- Li, Q., Ma, C., Tai, H., Qiu, H., and Yang, A. (2020a). Comparative transcriptome analysis of two rice genotypes differing in their tolerance to saline-alkaline stress. *PLoS One* 15:e0243112. doi: 10.1371/journal.pone.0243112
- Li, X., Zheng, H., Wu, W., Liu, H., Wang, J., Jia, Y., et al. (2020b). QTL mapping and candidate gene analysis for alkali tolerance in japonica Rice at the bud stage based on linkage mapping and genome-wide association study. *Rice* 13:48. doi: 10.1186/s12284-020-00412-5
- Liang, J. L., Qu, Y. P., Yang, C. G., Ma, X. D., Cao, G. L., Zhao, Z. W., et al. (2015). Identification of QTLs associated with salt or alkaline tolerance at the seedling stage in rice under salt or alkaline stress. *Euphytica* 201, 441–452. doi: 10.1007/s10681-014-1236-8
- Liang, X., Chen, X., Li, C., Fan, J., and Guo, Z. (2017). Metabolic and transcriptional alternations for defense by interfering OsWRKY62 and OsWRKY76 transcriptions in rice. *Sci. Rep.* 7:2474. doi: 10.1038/s41598-017-02643-x
- Livak, K. J., and Schmittgen, T. D. (2002). Analysis of relative gene expression data using real-time quantitative PCR. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262
- Lv, B. S., Li, X. W., Ma, H. Y., Sun, Y., Wei, L. X., Jiang, C. J., et al. (2013). Differences in growth and physiology of Rice in response to different saline-alkaline stress factors. *Agron. J.* 105, 1119–1128. doi: 10.2134/agronj2013.0017
- Ma, L., Qing, C., Zhang, M., Zou, C., Pan, G., and Shen, Y. (2021a). GWAS with a PCA uncovers candidate genes for accumulations of microelements in maize seedlings. *Physiol. Plant.* 172, 2170–2180. doi: 10.1111/ppl.13466
- Ma, L., Zhang, M., Chen, J., Qing, C., He, S., Zou, C., et al. (2021b). GWAS and WGCNA uncover hub genes controlling salt tolerance in maize (*Zea mays* L.) seedlings. *Theor. Appl. Genet.* 134, 3305–3318. doi: 10.1007/s00122-021-03897-w
- Munns, R., and Tester, M. (2008). Mechanisms of salinity tolerance. *Annu. Rev. Plant Biol.* 59, 651–681. doi: 10.1146/annurev.arplant.59.032607.092911
- Nakamura, A., Fukuda, A., Sakai, S., and Tanaka, Y. (2006). Molecular cloning, functional expression and subcellular localization of two putative vacuolar voltage-gated chloride channels in rice (*Oryza sativa* L.). *Plant Cell Physiol.* 47, 32–42. doi: 10.1093/pcp/pci220
- Naveed, S. A., Zhang, F., Zhang, J., Zheng, T.-Q., Meng, L.-J., Pang, Y.-L., et al. (2018). Identification of QTN and candidate genes for salinity tolerance at the germination and seedling stages in rice by genome-wide association analyses. *Sci. Rep.* 8:6505. doi: 10.1038/s41598-018-24946-3
- Ni, L., Wang, S., Shen, T., Wang, Q., Chen, C., Xia, J., et al. (2020). Calcium/calmodulin-dependent protein kinase OsDMI3 positively regulates saline-alkaline tolerance in rice roots. *Plant Signal. Behav.* 15:1813999. doi: 10.1080/15592324.2020.1813999
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qadir, M., Quillerou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R. J., et al. (2014). Economics of salt-induced land degradation and restoration. *Nat. Res. Forum* 38, 282–295. doi: 10.1111/1477-8947.12054
- Qi, D., Han, L., and Zhang, S. (2005). Methods of characterization and evaluation of salt or alkaline tolerance in rice. *J. Plant Genet. Resour.* 6, 226–231. doi: 10.3969/j.issn.1672-1810.2005.02.021
- Qi, D., Zhang, S., Cao, G., Ruan, R., Sun, M., Zhang, Y., et al. (2006). Studies on screening methods for alkaline tolerance at germination period and early seedling stage in Rice. *J. Plant Genet. Resour.* 7, 74–80. doi: 10.3969/j.issn.1672-1810.2006.01.015
- Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2013). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8:e66428. doi: 10.1371/journal.pone.0066428
- Ren, Z. H., Gao, J. P., Li, L. G., Cai, X. L., Huang, W., Chao, D. Y., et al. (2005). A rice quantitative trait locus for salt tolerance encodes a sodium transporter. *Nat. Genet.* 37, 1141–1146. doi: 10.1038/ng1643
- Ross, C. A., Liu, Y., and Shen, Q. J. (2007). The WRKY gene family in rice (*Oryza sativa*). *J. Integr. Plant Biol.* 49, 827–842. doi: 10.1111/j.1744-7909.2007.00504.x
- Shi, D., and Yin, L. (1993). Difference between salt (NaCl) and alkaline (Na₂CO₃) stresses on *Puccinellia tenuiflora* (Griseb.) Scribn. et Merr. plants. *J. Integr. Plant Biol.* 35, 144–149.
- Shi, Y. Y., Gao, L. L., Wu, Z. C., Zhang, X. J., Wang, M. M., Zhang, C. S., et al. (2017). Genome-wide association study of salt tolerance at the seed germination stage in rice. *BMC Plant Biol.* 17:92. doi: 10.1186/s12870-017-1044-0
- Sun, J., Wang, J., Guo, W., Yin, T., Zhang, S., Wang, L., et al. (2021). Identification of alkali-tolerant candidate genes using the NGS-assisted BSA strategy in rice. *Mol. Breed.* 41:44. doi: 10.1007/s11032-021-01228-x
- Tekeu, H., Ngonkeu, E. L. M., Belanger, S., Djougoue, P. F., Abed, A., Torkamaneh, D., et al. (2021). GWAS identifies an ortholog of the rice D11 gene as a candidate gene for grain size in an international collection of hexaploid wheat. *Sci. Rep.* 11:19483. doi: 10.1038/s41598-021-98626-0
- Thomson, M. J., de Ocampo, M., Egdane, J., Rahman, M. A., Sajise, A. G., Adorada, D. L., et al. (2010). Characterizing the *Saltol* quantitative trait locus for salinity tolerance in rice. *Rice* 3, 148–160. doi: 10.1007/s12284-010-9053-8
- Turner, S. D. (2014). Qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. *J. Open Source Softw.* 3:731. doi: 10.21105/joss.00731
- Valenzuela, C. E., Acevedo-Acevedo, O., Miranda, G. S., Vergara-Barros, P., Holuigue, L., Figueroa, C. R., et al. (2016). Salt stress response triggers activation of the jasmonate signaling pathway leading to inhibition of cell elongation in Arabidopsis primary root. *J. Exp. Bot.* 67, 4209–4220. doi: 10.1093/jxb/erw202

- Vengosh, A. (2014). "11.9: salinization and saline environments," in *Treatise on Geochemistry. 2nd Edn.* eds. H. D. Holland and K. K. Turekian (Oxford: Elsevier), 325–378.
- Wang, H., Lin, X., Cao, S., and Wu, Z. (2015). Alkali tolerance in rice (*Oryza sativa* L.): growth, photosynthesis, nitrogen metabolism, and ion homeostasis. *Photosynthetica* 53, 55–65. doi: 10.1007/s11099-015-0079-4
- Wang, W. S., Mauleon, R., Hu, Z. Q., Chebotarov, D., Tai, S. S., Wu, Z. C., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9
- Wang, Z., Wang, J., Bao, Y., Wu, Y., Su, X., and Zhang, H. (2010). Inheritance of rice seed germination ability under salt stress. *Rice Sci.* 17, 105–110. doi: 10.1016/S1672-6308(08)60112-2
- Xiong, H., Li, J., Liu, P., Duan, J., Zhao, Y., Guo, X., et al. (2014). Overexpression of OsMYB48-1, a novel MYB-related transcription factor, enhances drought and salinity tolerance in Rice. *PLoS One* 9:e92913. doi: 10.1371/journal.pone.0092913
- Yang, C. W., Shi, D. C., and Wang, D. L. (2008). Comparative effects of salt and alkali stresses on growth, osmotic adjustment and ionic balance of an alkali-resistant halophyte *Suaeda glauca* (Bge.). *Plant Growth Regul.* 56, 179–190. doi: 10.1007/s10725-008-9299-y
- Yang, J., Lee, S. H., Goddard, M. E., and Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* 88, 76–82. doi: 10.1016/j.ajhg.2010.11.011
- Yao, W., Li, G., Yu, Y., and Ouyang, Y. (2018). funRiceGenes dataset for comprehensive understanding and application of rice functional genes. *GigaScience* 7, 1–9. doi: 10.1093/gigascience/gix119
- Yu, J., Zao, W. G., He, Q., Kim, T. S., and Park, Y. J. (2017). Genome-wide association study and gene set analysis for understanding candidate genes involved in salt tolerance at the rice seedling stage. *Mol. Gen. Genomics.* 292, 1391–1403. doi: 10.1007/s00438-017-1354-9
- Yutaka, S., Hinako, T., Kaori, K., Hiroshi, M., Nobukazu, N., Hiroshi, I., et al. (2013). RiceXPro version 3.0: expanding the informatics resource for rice transcriptome. *Nucleic Acids Res.* 41, D1206–D1213. doi: 10.1093/nar/gks1125
- Zhang, F., Wang, C., Li, M., Cui, Y., Shi, Y., Wu, Z., et al. (2021). The landscape of gene-CDS-haplotype diversity in rice (*Oryza sativa* L.): properties, population organization, footprints of domestication and breeding, and implications in genetic improvement. *Mol. Plant* 14, 787–804. doi: 10.1016/j.molp.2021.02.003
- Zhang, H., Liu, X. L., Zhang, R. X., Yuan, H. Y., Wang, M. M., Yang, H. Y., et al. (2017). Root damage under alkaline stress is associated with reactive oxygen species accumulation in Rice (*Oryza sativa* L.). *Front. Plant Sci.* 8:1580. doi: 10.3389/fpls.2017.01580

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Mei, Zhang, Jiang, Lu 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) and the copyright owner(s) 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 and Application of Intragenic Markers for 14 Nitrogen-Use Efficiency Genes in Rice (*Oryza sativa* L.)

Pingbo Li¹, Zhen Li¹, Xu Liu¹, Hua Zhang¹, Qingguo Wang¹, Nana Li², Hanfeng Ding² and Fangyin Yao^{1*}

¹ Institute of Wetland Agriculture and Ecology, Shandong Academy of Agricultural Sciences, Jinan, China, ² Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences, Jinan, China

OPEN ACCESS

Edited by:

Xia Xin,
Institute of Crop Sciences (CAAS),
China

Reviewed by:

Weihua Qiao,
Institute of Crop Science (CAAS),
China
Jian Sun,
Shenyang Agricultural University,
China

*Correspondence:

Fangyin Yao
yaofy@163.com

Specialty section:

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

Received: 08 March 2022

Accepted: 28 March 2022

Published: 09 May 2022

Citation:

Li P, Li Z, Liu X, Zhang H,
Wang Q, Li N, Ding H and Yao F
(2022) Development and Application
of Intragenic Markers for 14
Nitrogen-Use Efficiency Genes in Rice
(*Oryza sativa* L.).
Front. Plant Sci. 13:891860.
doi: 10.3389/fpls.2022.891860

Asian cultivated rice consists of two main subspecies, *xian/indica* (XI) and *geng/japonica* (GJ), and GJ accessions have significantly lower nitrogen-use efficiency (NUE) than XI accessions. In order to facilitate genetic improvement of NUE in GJ accessions, we conducted haplotype analysis of 14 cloned NUE genes using 36 rice germplasm accessions with high-quality reference genome and developed 18 intragenic markers for elite haplotypes, which were then used to evaluate NUE genes in another 41 genetically diverse germplasm accessions from 12 countries and 71 approved GJ cultivars from northern provinces of China. Our results show that elite haplotypes of 12 NUE genes are mainly existed in XI accessions, but few is distributed in GJ accessions. The number of elite haplotypes carried by an XI accession can reach 10, while that carried by a GJ accession is less than 3. Surprisingly, the elite haplotype of gene *DEP1* is nearly fixed in approved GJ cultivars, and elite haplotypes of gene *MYB61* and *NGR5* have been introduced into some approved GJ cultivars. The developed intragenic markers for NUE genes and evaluated 77 genetically diverse rice accessions could be of great use in the improvement of NUE in GJ cultivars.

Keywords: rice, nitrogen-use efficiency, haplotype analysis, intragenic marker, germplasm accession

INTRODUCTION

Rice is a staple food that feeds more than half of the world population. Nitrogen is a macronutrient that is essential to the growth and grain yield of rice plants. The adoption of semidwarf rice varieties and nitrogen fertilizer has strikingly boosted rice yield since 1960s (Khush, 1999). However, the excess input of nitrogen fertilizer not only contributes little to the increase in grain yield, but also brings great threat to environment in recent years. Therefore, adoption of rice cultivars with high nitrogen-use efficiency (NUE) and reduction of nitrogen input is a promising way to achieve sustainable agriculture. Asian cultivated rice consists of two main subspecies, *xian/indica* (XI) and *geng/japonica* (GJ), which display significant difference in various traits, including NUE (Hu et al., 2015; Gao et al., 2019). GJ cultivars, favored by consumers in Japan, South Korea, and the northern provinces of China, have low NUE, and massive input of nitrogen fertilizer is the way selected by farmers to improve rice yield. Therefore, genetic improvement of NUE in GJ cultivars is of great necessity to achieve sustainable agriculture. NUE involves nitrogen uptake, translocation, and

assimilation, and results in changes in plant height, tiller number, and grain yield in rice (Li et al., 2017). Thus, it is a complex trait controlled by quantitative trait loci (QTL) (Zhou et al., 2017; Tang et al., 2019; Shen et al., 2021). Up to now, 14 genes conferring natural variation in NUE have been cloned and characterized (Table 1). Introgression of elite alleles or haplotypes of these cloned genes into the background of GJ cultivars could be a promising way to improve NUE of GJ cultivars. Marker-assisted selection has been proved to be an effective method in genetic improvement of rice cultivars, of which the premise is the development of effective molecular markers for target genes (Wang et al., 2017; Jiang et al., 2019; Mao et al., 2021).

Objective

Our aim was to develop effective intragenic markers for the 14 cloned NUE genes in rice, and then to evaluate NUE genes of some germplasm accessions and approved GJ cultivars, which could guide the following genetic improvement. Haplotype analysis of the 14 NUE genes from 36 rice accessions with high-quality reference genome provided many useful sequence variations for marker development, and developed intragenic markers facilitated the evaluation of NUE genes in germplasm accessions and approved cultivars.

MATERIALS AND METHODS

Rice Accessions and Cultivars

Three panels of rice materials were exploited in this study. The first panel consisted of 36 rice accessions with reference genome (Table 2), including GJ accession Nipponbare (NIP) (International Rice Genome Sequencing Project, 2005), 31 genetically diverse rice accessions (Qin et al., 2021), XI accessions Huazhan and Tianfeng (Zhang et al., 2022), and XI accessions Minghui63 and Zhenshan97 (Song et al., 2021). The second panel consisted of 41 genetically diverse rice accessions from 12 countries (Supplementary Table S16). The third panel consisted of 71 GJ cultivars approved in the northern provinces of China, including 2 from Anhui, 1 from Henan, 9 from Jiangsu, 39 from Shandong, 1 from Tianjin, and 19 approved in several provinces (Supplementary Table S17). The subpopulation of accessions in the first panel and second panel referred to a previous study (Wang W. et al., 2018).

Haplotype Analysis

The genomic sequences of 14 cloned NUE genes from 36 rice accessions with reference genome were extracted using the software BioEdit (Hall, 1999) and software SeqBuilder from the Lasergene package (DNASTAR, Inc., Madison, USA), including 2-kb region upstream of the start codon termed as 5' untranslated region (5'UTR), the coding region, and 1-kb region downstream of the stop codon termed as 3'UTR. Elite alleles of *NGR2*, *NGR5*, *DEP1*, and *IR36* were reported to be carried by a very few accessions not included in the 36 accessions, and their genomic sequences could not be found in related databases and papers. Therefore, genomic sequences of allele *ngr2* from accession NM73, allele *NGR5* from accession Guichao2, allele *dep1* from accession Ballila, and allele *NAC42* from accession IR36 were

resequenced using the Sanger sequencing method and assembled using the software SeqMan from the Lasergene package, and related primers were listed in Supplementary Table S1. Sequence alignment for each gene was conducted using the software MEGA 7. Haplotype analysis of each gene was conducted, based on a combination of functional variations reported in previous studies and sequence variations identified in this study. For each NUE gene, the haplotype of accession NIP was defined as HapA, and reported elite haplotype was defined as HapB (Table 2 and Supplementary Tables S2–S15).

Marker Development

Several rules were followed during marker development. First of all, target variations should be unique to the elite haplotype of each gene. Secondly, an insertion/deletion (Indel) whose size was between 5 and 30 bp was given priority, and the Indel marker with amplification fragment between 100 and 300 bp was developed. Thirdly, if no Indel could be selected, a single-nucleotide polymorphism (SNP) that could be transformed into a common restriction enzyme site was taken into consideration, and then the cleaved amplified polymorphic sequences (CAPS) marker with amplification fragment between 100 and 500 bp or derived CAPS (dCAPS) marker with amplification fragment between 100 and 300 bp was developed. Finally, if no variation could meet the second and third rules, a penta-primer amplification-refractory mutation system (PARMS) marker was designed (Qing et al., 2018).

Indel markers were developed using the software Primer Premier 6 (PREMIER Biosoft, San Francisco, USA). CAPS markers and dCAPS markers were developed using the website dCAPS Finder 2.0¹ (Neff et al., 2002) and the software Primer Premier 6. PARMS markers were developed using the website SNPWay².

Marker Analysis

For Indel markers, CAPS markers and dCAPS markers, the PCR reaction was 3-μl genomic DNA with a concentration of 20 ng/μl, 0.5 μl of each primer with a concentration of 10 μM/L, 10-μl 2**Taq* PCR MasterMix, and 6 μl of ddH₂O. The PCR profile was 3 min at 94°C for denaturation, followed by 32 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min/kb, then by 3 min at 72°C for extension, and finally by 1 min at room temperature. The PCR products of Indel markers were run on 4% polyacrylamide gels with a voltage of 160 V for 80 min, and bands were then revealed using a silver staining procedure: gels were washed for 1 min with dH₂O, followed by incubated for 10 min in staining solution (0.4 g AgNO₃, 400 ml dH₂O) with gentle shaking, then washed two times for 1 min with dH₂O, then incubated for 4 min in developing solution (6 g NaOH, 2 ml CH₃CHO, 400 ml dH₂O) with gentle shaking, and finally washed for 1 min with water. The PCR products of dCAPS markers and CAPS markers were digested with corresponding restriction enzymes purchased from Takara Biomedical Technology Co., LTD., Beijing, China, according to the recommended reaction and temperature. The digested products of dCAPS markers were

¹<http://helix.wustl.edu/dcaps/dcaps.html>

²<http://www.snpway.com/>

TABLE 1 | Information of 14 cloned NUE genes.

Gene symbol	LOC number	Protein encoding ^a	Subcellular localization	Protein function	Elite allele or haplotype	Reference
<i>OsNPF6.1</i>	LOC_Os01g01360	A nitrate transporter	Plasma membrane	Uptake and redistribution of nitrate	<i>OsNPF6.1^{HapB}</i> that carries two novel binding sites of OsNAC42	Tang et al., 2019
<i>DNR1</i>	LOC_Os01g08270	An aminotransferase	Nucleus, cytoplasm	Catalyze the conversion of indole-3-pyruvate to L-Trp, thus antagonizing auxin biosynthesis	Hap.A showing low mRNA abundance	Zhang et al., 2021
<i>MYB61</i>	LOC_Os01g18240	A MYB family TF	Nucleus		Hap1 that carries the deletion of a helitron element in the 5'UTR and thus facilitates the binding of NGR2/OsGRF4	Gao et al., 2020
<i>SBM1</i>	LOC_Os01g65120	An oligopeptide transporter	Plasma membrane	Promote nitrogen uptake and transport	HapB carried by accession Kasalath and PA64S	Xu et al., 2021
<i>NGR2</i>	LOC_Os02g47280	A GRF family TF	Nucleus	Transcriptional activation of nitrogen metabolism related genes	Hap.B or <i>ngr2</i> that is carried by accession NM73 and shows high mRNA abundance	Li et al., 2018
<i>OsNR2</i>	LOC_Os02g53130	A NADH/NADPH-dependent nitrate reductase		Promote nitrate uptake and utilization	Allele that carries a SNP conferring the Trp ₇₇₉ to Arg ₇₈₃ substitution in the NAD(P) binding domain	Gao et al., 2019
<i>NGR5</i>	LOC_Os05g32270	An AP2-domain TF	Nucleus	Facilitate nitrogen-dependent recruitment of polycomb repressive complex 2 to repress branching-inhibitory genes <i>via</i> H3K27me3 modification	Hap.2 that is carried by accession Guichao2 and shows high mRNA abundance	Wu et al., 2020
<i>OsTCP19</i>	LOC_Os06g12230	A TCP family TF	Nucleus	Transcriptional repression of DLT genes	<i>OsTCP19-H</i> that carries a 29-bp deletion in the promoter and thus enhances the <i>trans</i> -repression of LBD proteins	Liu et al., 2021
<i>ARE1</i>	LOC_Os08g12780	A protein with unknown function	Chloroplast		<i>ARE1⁹³¹¹</i> and <i>ARE1^{MH63}</i> that carry small insertions in the promoter and show low mRNA abundance	Wang W. et al., 2018
<i>DEP1</i>	LOC_Os09g26999	An atypical γ subunit of G proteins	Nucleus		<i>dep1</i> that is carried by accession Ballila and carries the replacement of a 637-bp stretch of the middle of exon 5 by a 12-bp sequence	Huang et al., 2009; Sun et al., 2014
<i>OsNAC42</i>	LOC_Os09g32040	A NAC family TF	Nucleus	Transcriptional activation of <i>OsNPF6.1</i>	<i>OsNAC42^{HapC}</i> that is carried by accession IR36 and shows high mRNA abundance	Tang et al., 2019
<i>OsNLP4</i>	LOC_Os09g37710	An NLP family TF	Nucleus	Transcriptional activation of <i>OsNIR</i>	<i>OsNLP4^{HapB}</i> that shows high mRNA abundance	Yu et al., 2021
<i>NRT1.1B</i>	LOC_Os10g40600	A nitrate transporter	Plasma membrane	Uptake and transport of nitrate	Allele that carries a SNP conferring the Thr ₃₂₇ to Met ₃₂₇ substitution in the central cytoplasmic loop	Hu et al., 2015
<i>TOND1</i>	LOC_Os12g43440	A thaumatin protein	Plasma membrane		H1 carried by accession Teqing and 9311	Zhang et al., 2015

^aTF, transcription factor.

TABLE 2 | Haplotypes of 14 NUE genes in the 36 rice accessions with reference genome.

Accessions	Origin ^a	Subpopu. ^b	NUE genes ^c														No. of elite haplotypes
			<i>OsNPF6.1</i>	<i>DNR1</i>	<i>MYB61</i>	<i>SBM1</i>	<i>NGR2</i>	<i>OsNR2</i>	<i>NGR5</i>	<i>OsTCP19</i>	<i>ARE1</i>	<i>DEP1</i>	<i>NAC42</i>	<i>OsNLP4</i>	<i>NRT1.1B</i>	<i>TOND1</i>	
N22	India	cA	HapD	n.d.	n.d.	HapB	HapE	HapB	HapD	HapB	n.d.	HapF	HapD	HapB	n.d.	n.d.	4
Basmati 1	Pakistan	cB	HapA	HapC	HapB	HapA	HapE	n.d.	HapA	n.d.	n.d.	HapA	HapE	HapA	HapA	n.d.	1
LJ	YN, China	GJ-adm	HapA	HapD	HapC	HapA	HapE	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
NamRoo	Thailand	GJ-sbtrp	HapA	HapA	HapC	HapA	HapE	HapA	HapA	HapA	n.d.	HapA	HapA	HapA	HapA	HapC	0
2428	JS, China	GJ-tmp	HapA	HapD	HapC	HapA	HapC	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
DHX2	HLJ, China	GJ-tmp	HapA	HapA	HapC	HapA	HapE	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
Kosh	Japan	GJ-tmp	HapA	HapA	HapA	HapA	HapC	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
KY131	Japan	GJ-tmp	HapA	HapA	HapC	HapA	HapC	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
NIP	Japan	GJ-tmp	HapA	HapA	HapA	HapA	HapC	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	0
ZH11	TJ, China	GJ-tmp	HapA	HapA	HapC	HapA	HapC	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
Lemont	United States	GJ-trp	HapA	HapD	HapC	HapD	HapE	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapA	HapC	0
HZ	China	XI	HapA	HapB	HapB	HapD	HapA	HapB	HapD	HapB	HapC	HapC	HapC	HapB	HapB	HapB	7
TFB	GD, China	XI	HapB	HapB	HapB	HapC	HapB	HapB	HapD	HapA	HapC	HapF	HapA	HapB	HapB	HapC	8
CN1	SC, China	XI-1A	HapB	HapB	HapB	HapC	HapD	HapB	HapE	HapA	HapC	HapF	HapE	HapB	HapB	HapB	8
D62	SC, China	XI-1A	HapB	HapB	HapB	HapD	HapC	HapB	HapD	HapA	HapC	HapF	HapB	HapB	HapB	HapB	9
DG	FJ, China	XI-1A	HapC	HapB	HapB	HapC	HapD	HapB	HapD	HapA	HapC	HapD	HapD	HapB	HapB	HapC	6
FS32	n.d.	XI-1A	HapC	HapB	HapB	n.d.	HapA	HapB	HapE	HapA	HapC	HapD	n.d.	HapB	HapB	HapB	7
G46	SC, China	XI-1A	HapB	HapB	HapB	n.d.	HapA	HapB	HapE	HapA	HapC	HapF	HapE	HapB	HapB	HapC	7
II32	HN, China	XI-1A	HapB	HapB	HapB	HapC	HapB	HapB	HapE	HapA	HapC	HapD	HapA	HapB	HapB	HapC	8
ZS97	ZJ, China	XI-1A	HapB	HapB	HapB	HapC	HapB	HapB	HapE	HapA	HapC	HapD	HapA	HapB	HapB	HapC	8
9311	JS, China	XI-1B	HapA	HapB	HapB	HapC	HapA	HapB	HapE	HapA	HapC	HapD	HapD	HapB	HapB	HapB	6
G8	GD, China	XI-1B	HapA	HapB	HapB	HapC	HapB	HapB	HapE	HapA	HapC	HapC	HapD	HapB	HapB	HapB	8
IR64	IRRI	XI-1B	HapA	HapB	HapB	HapD	HapB	HapB	HapC	HapB	HapB	HapE	HapB	HapB	HapB	HapB	10
J4115	HN, China	XI-1B	HapA	HapB	HapB	HapD	HapE	HapB	HapB	HapA	HapB	HapC	HapB	HapB	HapB	HapB	9
R498	SC, China	XI-1B	HapC	HapB	HapB	HapD	HapA	HapB	HapB	HapA	HapB	HapC	HapB	HapB	HapB	HapB	9
R527	JX China	XI-1B	HapA	HapB	HapB	HapD	HapB	HapB	HapB	HapA	HapB	HapE	HapB	HapB	HapB	HapB	10
S548	SC, China	XI-1B	HapA	HapB	HapB	HapD	HapA	HapA	HapB	HapB	HapB	HapC	HapC	HapB	HapB	HapB	8
Y3551	SC, China	XI-1B	HapA	HapB	HapB	HapD	HapA	HapB	HapB	HapA	HapC	HapC	HapE	HapB	HapB	HapB	8
Y58S	HN, China	XI-1B	HapA	HapB	HapB	HapD	HapE	HapB	HapA	HapA	HapC	HapC	HapB	HapB	HapB	HapB	7
TM	Madagascar	XI-2	HapA	HapB	n.d.	HapC	HapE	HapB	HapD	HapB	n.d.	HapD	HapD	HapB	HapB	HapB	6
TUMBA	Indonesia	XI-3	HapA	HapB	HapB	HapC	HapD	HapB	HapD	HapB	n.d.	HapE	HapB	HapB	HapB	HapB	8
FH838	SC, China	XI-adm	HapA	HapB	HapB	HapD	HapA	HapB	HapB	HapA	HapB	HapE	HapB	HapB	HapB	HapC	8
G630	Guyana	XI-adm	HapA	HapB	HapB	HapD	HapA	HapB	HapB	HapA	n.d.	HapC	HapE	HapB	HapB	HapC	6
MH63	FJ, China	XI-adm	HapA	HapB	HapB	HapD	HapA	HapB	HapB	HapA	HapB	HapC	HapE	HapB	HapB	HapB	8
WSSM	GD, China	XI-adm	HapA	HapB	HapB	HapD	HapA	HapB	HapD	HapA	HapC	HapE	HapB	HapB	HapB	n.d.	7
YX1	SC, China	XI-adm	HapB	HapB	HapB	HapC	HapB	HapB	HapB	HapA	HapC	HapD	HapA	HapB	HapB	HapB	10

^aFJ, Fujian; GD, Guangdong; HLJ, Heilongjiang; HN, Hunan; JS, Jiangsu; JX, Jiangxi; SC, Sichuan; TJ, Tianjin; YN, Yunnan; ZJ, Zhejiang. ^bThe subpopulation information referred to the paper reported by Qin et al. (2021). ^cn.d., not determined. If the genomic sequence of a gene from an accession could not be extracted, or carried many novel variations, the haplotype of the gene would be termed as n.d.

TABLE 3 | Intragenic markers for the 14 NUE genes.

Gene	Variation position ^a	Marker type	Marker name ^b	Primer sequence	Reference size ^c	Restriction enzyme	Target Haplotype
<i>OsNPF6.1</i>	-1197	Indel	NPF6.1 5U ID	F TAATGTCCTTTCCCGTGTT R GTACTACTTCGGCTGTCC	220 (+20)		HapB
<i>DNR1</i>	+2165	Indel	DNR1 I4 ID	F CGTCAATTATGGTTACCTCTG R GCATCTCATAGAACTGAAGAAG	171 (-10)		HapB
<i>MYB61</i>	+2346	Indel	MYB61 3U ID	F ACTTGAATACAGGCATGGAA R CGTTATGCTTGTGCTTGA	205 (+26)		HapB
<i>SBM1</i>	-325	Indel	SBM1 5U ID	F TCGGGTGACTACGGATC R GCACATACATATCAGGGA	90 (-8)		HapB
	+865	CAPS	SBM1 E3 S	F TCTTCAACTGGATCAACTTC R TACATCACGGTGGTCATC	380/(100,280)	<i>Bgl</i> I	HapB, HapD
<i>NGR2</i>	-935	CAPS	NGR2 5U S	F TCATTGACCTACGGTTGC R GCTGCTCCAACATCTTCT	(249,24)/273	<i>Taq</i> I	HapB
	+1701	PARMS	NGR2 I3 S	FC GAAGGTGACCAAGTTCATGCTAGTAGTACTACTTCGATTTGGTGCT FT GAAGGTGCGGAGTCAACGGATTAGTAGTACTACTTCGATTTGGTGCC R GCAGTGCAGAGAGTAAAGTTTCAG			HapB
<i>'OsNR2</i>	+2202	Indel	NR2 E4 ID	F TCACGTCCATCGTTGAGA R ACAGGCTCTTCTGTGCCAT	120 (+12)		HapB
<i>NGR5</i>	-1565	Indel	NGR5 5U ID	F AGAACACACGGGATAGGAT R GAATCACTTGCTCGCTAGA	210 (-13)		HapB, HapC, HapD
	+3326	PARMS	NGR5 I6 S	FC GAAGGTGACCAAGTTCATGCTATCTCGCTGTCCTAAACGACTTCC FT GAAGGTGCGGAGTCAACGGATTATCTCGCTGTCCTAAACGACTTCT R GAATTACGTACACCCCTCCGTACTC			HapB
<i>OsTCP19</i>	-2022	Indel	TCP19 5U ID	F AACTCTTCAGGGTTCCTTGC R GTGCCGTGTCACATAGAG	231 (-29)		HapB
<i>ARE1</i>	-393	Indel	ARE1 5U ID	F CCGTGTCTTATCCACTCC R ATGGGGATCGATACGATG	115 (+6)		HapB
	-723	CAPS	ARE1 5U S	F TTAACACTTGTGGCAATGAC R CTAGTACCGTATTGGCTGTT	320/(100,210)	<i>Ddel</i>	HapC
<i>DEP1</i>	+3454	Indel	DEP1 E5 ID	F GACCAAGGTGCCTCAATT R TTCAACCTCGTCTCATAGC	1085 (-615)		HapB

(Continued)

TABLE 3 | (Continued)

Gene	Variation position ^a	Marker type	Marker name ^b	Primer sequence	Reference size ^c	Restriction enzyme	Target Haplotype
NAC42	+3508	dCAPS	NAC42 3U S	F TACGTGACTTCGACGGCIGA R ACGGTCCAAATGCTGCTTCG	(100,20)/120	DdeI	HapB
NLP4	-1771	Indel	NLP4 5U ID	F AAGTCCTTCTTAAACTGAGA R GGTCTGTTCCAAACAAAGAT	136 (-6)		HapB
NRT1.1B	+2825	Indel	NRT1.1B 11 ID	F CATATTTGTTGGCTGCTAAC R GGTGGTTCTAACGGTCAA	178 (+16)		HapB
TOND1	-1205	Indel	TOND1 5U ID	F TTGGGTCCCTGACAAATA R TGAACAACCTCAAGTAGCA	153 (+8)		HapB

^aVariation position is the position of target variation selected for marker development, relative to the start codon ATG. Target variation and its position were also shown in the supplementary table of each gene.
^bThe name of each marker was defined as "gene name + variation position + variation type." For example, "NPF6.1 5U ID" indicates that the target variation is an Indel in the 5'UTR of NPF6.1, and "SBM1 E3 S" indicates that the target variation is an SNP in the third exon of SBM1.
^cFor Indel markers, the number out of bracket is the size of amplification fragment from accession NIP, and the content in bracket is the information of target Indel carried by the elite haplotype. For CAPS and dCAPS markers, the number before "/" is the size of fragment digested with corresponding restriction enzyme from NIP, while the number behind "/" is that from accessions carrying elite haplotype.

analyzed as that of Indel markers described above. The digested products of CAPS markers were run 2% agarose gels containing GeneRed nucleic acid dye with a voltage of 120 V for 20 min, and bands were developed in UVP EC3 Imaging System (Analytik Jena AG, Jena, Germany).

For PARMS markers, the PCR reaction was 2-μl genomic DNA with a concentration of 20 ng/μl, 0.15 μl of each forward primer and 0.4 μl of common reverse primer with a concentration of 10 μM/L, 5-μl 2*PARMS master mix containing two universal fluorescent primers, and 2.3 μl of ddH₂O. The PCR profile was 15 min at 94°C for denaturation, followed by 10 cycles of 94°C for 20 s and 65°C (with each cycle minus 0.8°C) for 1 min, then by 28 cycles of 94°C for 20 s and 57°C for 1 min, and finally by 1 min at temperature. The PCR products were analyzed using ABI QuantStudio 5 fluorescent quantitative PCR system (Applied Biosystem, Foster, USA).

RESULTS

Haplotype Analysis of 14 Nitrogen-Use Efficiency Genes

In order to have a closer examination of sequence variations in the 14 cloned NUE genes (Table 1), we extracted the genomic sequences from the 36 rice accessions with reference genome, including 2-kb 5'UTR, the coding region, and 1-kb 3'UTR. We also sequenced the genomic sequence of allele *ngr2* from accession NM73, allele *NGR5* from accession Guichao2, allele *dep1* from accession Ballila, and allele *NAC42* from accession IR36, which were reported elite alleles and whose sequences were not accessible. Subsequently, we performed sequence alignment and haplotype analysis of each gene, with the type of accession NIP termed as HapA and reported elite haplotype termed as HapB (Table 2).

For *OsNPF6.1*, four haplotypes were identified, including the two reported by Tang et al. (2019; Table 2 and Supplementary Table S2). Compared to HapA, both HapB and HapC carried two SNPs at positions -1041 and -443 that created two new binding sites for protein OsNAC42. In addition, HapB carried a non-synonymous SNP at position +125 and HapC carried two non-synonymous SNPs at positions +506 and +535 in the first exon.

For *DNR1*, four haplotypes were identified, including the three types reported by Zhang et al. (2021), which were HapA (Hap.B-TEJ), HapB (Hap.A), HapC, and HapD (Hap.B-TRJ) (Table 2 and Supplementary Table S3). The four types were distributed in GJ-tmp, XI, cA/cB, and GJ-trp accessions, respectively (Table 2). Compared to HapA, HapB carried a 520-bp deletion, a 2-bp deletion and twenty SNPs in 5'UTR, five Indels and eighteen SNPs in the coding region, and three SNPs in the 3'UTR. In addition, both HapC and HapD carried the 520-bp deletion at position -1201.

For *MYB61*, three haplotypes were identified, in agreement with the three types reported by Gao et al. (2020), which were HapA (Hap3), HapB (Hap1), and HapC (Hap2) (Table 2 and Supplementary Table S4). HapC differed from HapA by an SNP G/A at position -323, of which both were distributed in GJ accessions. HapB, the type distributed in XI accessions, carried

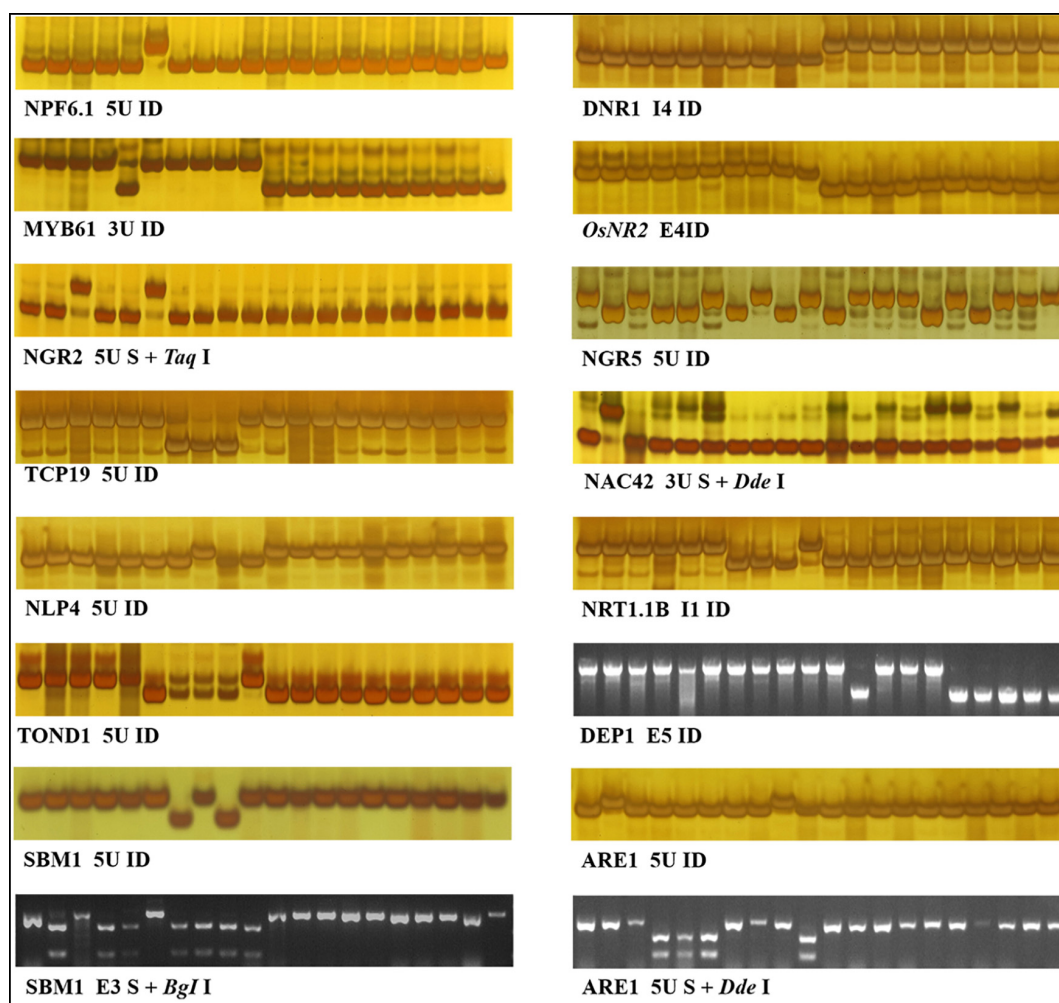


FIGURE 1 | Indel, CAPS, and dCAPS markers for 14 NUE genes. The 20 accessions used for marker evaluation were 9311, IR36, NM73, Guicao 2, Huagengxian 74, ZS97, Kasalath, Basmati 370, Dular, NJ11, 02428, Ballila, NIP, Kosh, KY131, Runnong11, Suxiu 867, Huaidao 5, Xudao 3, and Huageng5 (here from left to right). The first 15 accessions were germplasm accessions, including donors of the 14 NUE genes, and the other 5 accessions were mainstay cultivars in the Huang–Huai rice area of China. In the figure of marker “TOND1 5U ID,” no band was amplified from accession Kasalath, Basmati 370, and Dular.

the deletion of 975-bp helitron at position –446 that was close to the binding site of NGR2/OsGRF4 in the 5’UTR, and a non-synonymous SNP at position + 1224 in the coding region relative to the other two types.

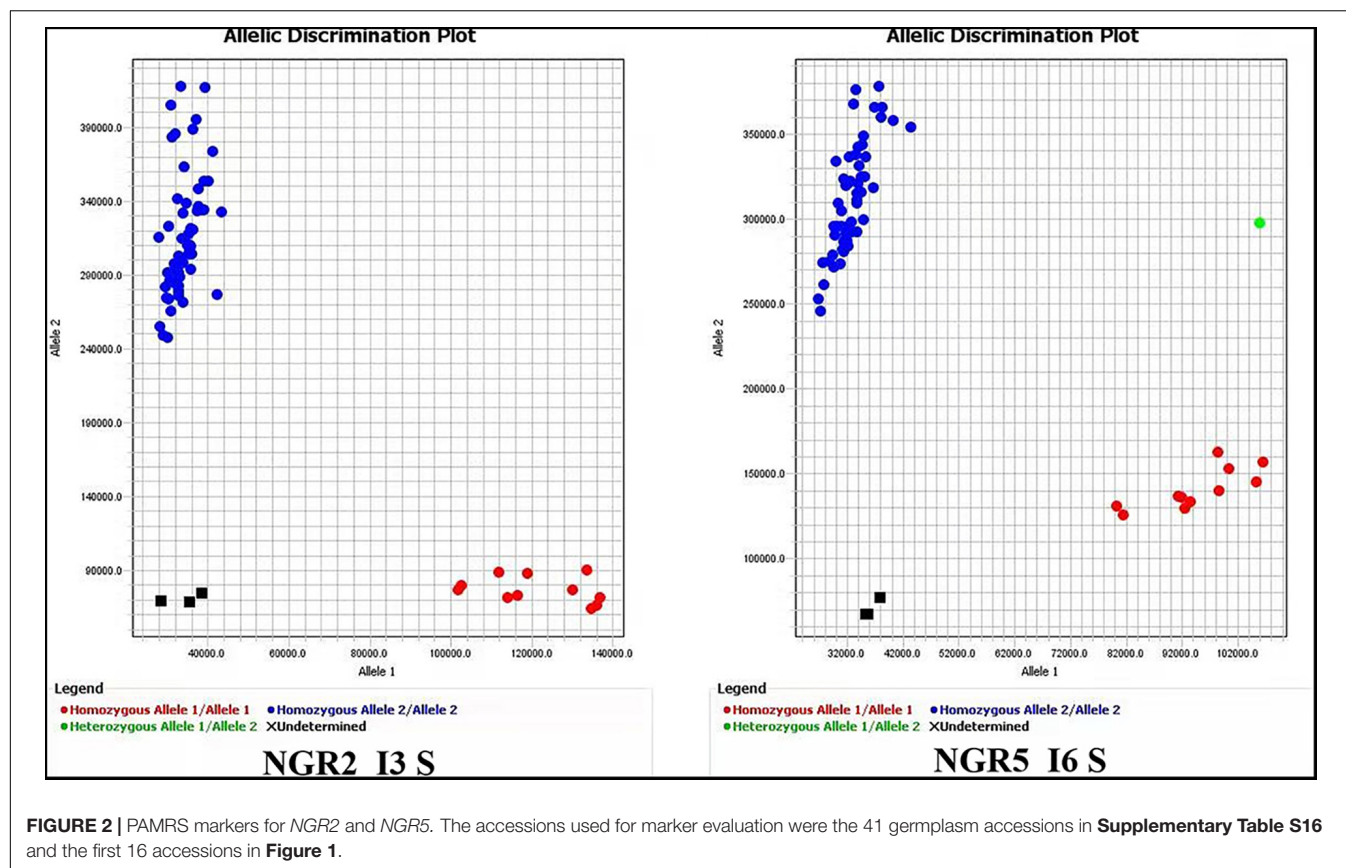
For *SBM1*, four haplotypes were identified, including the three types reported by Xu et al. (2021; **Table 2** and **Supplementary Table S5**). HapA and HapB were distributed mainly in GJ and cA accessions, respectively, while both HapC and HapD were distributed mainly in XI accessions. Compared to HapA and HapC, HapB carried two Indels and fifteen SNPs in the 5’UTR, a 3-bp deletion and two SNPs in the coding region, and three SNPs in the 3’UTR. HapD carries a 1,204-bp deletion in the 5’UTR and a non-synonymous SNP at position + 2 which disrupted the translation start site, relative to the other three types (**Supplementary Figure S1**).

For *NGR2*, five haplotypes were identified, including the three reported by Li et al. (2018) (**Table 2** and **Supplementary Table S6**). HapB carried 14 SNPs at positions –1806, –1754,

–1738, –1691, –1314, –935, –878, –844, –795, +511, +641, +654, +2275, and +3330 relative to the other four types, and was distributed in XI accessions. In addition, the two rare SNPs (+1187T > A and +1188C > A) preventing miR396-mediated cleavage of mRNA were not carried by all the 36 accessions.

For *OsNR2*, two haplotypes were identified, based on the functional SNP at position + 2692 reported by Gao et al. (2019; **Table 2** and **Supplementary Table S7**). HapA and HapB were distributed in GJ and XI accessions, respectively. HapB differed from HapA by a 26-bp insertion and nine SNPs in the 5’UTR, a 12-bp insertion and three non-synonymous SNPs in the coding region, and four SNPs in the 3’UTR.

For *NGR5*, five haplotypes were identified, in agreement with the five reported by Wu et al. (2020), which were HapA (Hap1), HapB (Hap2), HapC (Hap3), HapD (Hap4), and HapE (Hap5) (**Table 2** and **Supplementary Table S8**). Among those, HapB, HapC, and HapD shared many variations relative to the other two types, and differentiated from each other by an SNP, respectively.



HapB carried an SNP at position + 3326 in the sixth intron, and HapC carried an SNP at position + 4020 in the eighth exon.

For *OstCP19*, two haplotypes were identified, based on the functional 29-bp deletion at position −2022 reported by Liu et al. (2021; **Table 2** and **Supplementary Table S9**).

For *ARE1*, three haplotypes were identified, in agreement with the three reported by Wang Q. et al. (2018), which were HapA (*ARE1^{NPB}*), HapB (*ARE1^{MH63}*), and HapC (*ARE1⁹³¹¹*) (**Table 2** and **Supplementary Table S10**). Compared to HapA, HapB carried a 6-bp insertion and 10 SNPs in the 5'UTR, 11 SNPs in the coding region, and a 6-bp insertion and three SNPs in the 3'UTR. HapC carried a 2-bp deletion, a 7,808-bp insertion and eight SNPs in the 5'UTR, ten SNPs in the coding region, and a 6-bp insertion in the 3'UTR.

For *DEP1*, six haplotypes were identified (**Table 2** and **Supplementary Table S11**). The functional variation – the replacement of a 637-bp segment of the middle of exon 5 by a 12-bp sequence, is carried by accession Ballila but not by all 36 accessions (Huang et al., 2009).

For *OsNAC42*, five haplotypes were identified (**Table 2** and **Supplementary Table S12**). Compared to HapA, HapB carried eight SNPs in the 5'UTR, eleven SNPs in the coding region, and a 27-bp deletion and six SNPs in the 3'UTR.

For *OsNLP4*, two haplotypes were identified, in agreement with the two reported by Yu et al. (2021), which were distributed in GJ and XI accessions, respectively (**Table 2** and **Supplementary Table S13**). Compared to HapA, HapB carried

a 6-bp deletion and eight SNPs in the 5'UTR, five SNPs in the coding region, and five SNPs in the 3'UTR.

For *NRT1.1B*, two haplotypes were identified, based on the functional SNP at position + 3019 reported by Hu et al. (2015), which were distributed in GJ and XI accessions, respectively (**Table 2** and **Supplementary Table S14**). Compared to HapA, HapB carried a 3-bp insertion and thirteen SNPs in the 5'UTR, five Indels and twenty-two SNPs in the coding region, and two Indels and nine SNPs in the 3'UTR.

For *TOND1*, three haplotypes were identified, including the elite type reported by Zhang et al. (2015), which was HapB (H1) (**Table 2** and **Supplementary Table S15**). Compared to HapA and HapC, HapB carried an 8-bp insertion and three SNPs in the 5'UTR, and two non-synonymous SNPs in the coding region.

Development of Intragenic Markers for Nitrogen-Use Efficiency Genes

In order to develop effective markers for NUE genes, we focused on the variations only carried by HapB of each gene. As shown in **Table 3**, a total of 18 markers were developed for the 14 genes including twelve Indel markers, three CAPS makers, a dCAPS marker, and two PARMS markers. The effectiveness of 16 markers was evaluated by 20 germplasm accessions (**Figure 1**), and that of the two PARMS markers was evaluated by 57 germplasm accessions (**Figure 2**).

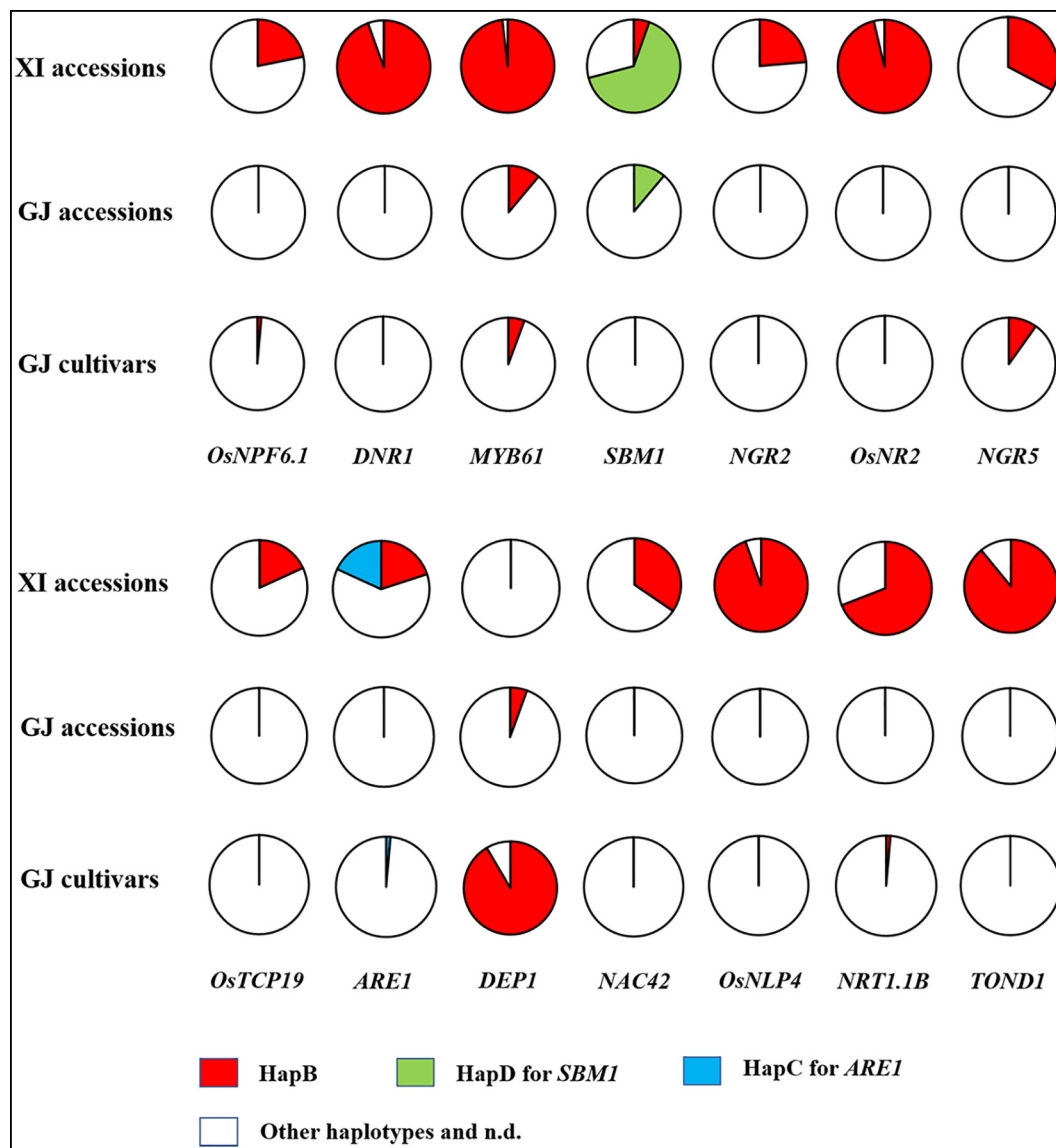


FIGURE 3 | Distribution of elite haplotypes of 14 NUE genes in 55 XI accessions, 18 GJ accessions, and 71 GJ cultivars.

SBM1 was reported as a negative regulator of NUE and grain yield, and the elite type HapB carried by accession Kasalath was a type with weak function (Xu et al., 2021). HapD carried a 1,204-bp deletion in the 5'UTR region and a non-synonymous SNP at position + 2 that would disrupt the translation start site, and was likely to have a better performance than HapB (Supplementary Table S5). Thus, the CAPS marker "SBM1 E3 S" together with restriction enzyme *Bgl* I was developed to select HapB and HapD, which could also differentiate the two types together with the Indel marker "SBM1 5U ID."

Two elite types of *ARE1* were reported, which were carried by accession Minghui63 and 9311, respectively (Wang Q. et al., 2018). However, the 6-bp insertion reported in the promoter region of accession 9311 was not observed in all the 36

accessions with reference genome, including the resequenced accession 9311. A 7,808-bp insertion in the HapC carried by 15 accessions was observed at position -466, which might reduce the expression of *ARE1* and make it an elite type (Supplementary Table S10). Thus, the CAPS marker "ARE1 5U S" together with restriction enzyme *Dde* I was developed to select HapC.

NGR5 was classified into five haplotypes, among which HapB, HapC, and HapD differentiated from each other only by one SNP (Supplementary Table S8). HapB was reported to be superior to HapC and HapD, but the SNP in the sixth intron could not account for the difference. We then examined the 7-kb region upstream of the start codon, and found that no additional variation could further differentiate the three types (data not shown). Therefore, it was likely that all the three types were elite,

and an Indel marker “NGR5 5U ID” was developed for the 13-bp deletion at position –1565, which was shared by the three types. In addition, a PARMS marker “NGR5 In6 S” was developed to select HapB only.

Evaluation of Nitrogen-Use Efficiency Genes in Germplasm Accessions and Approved Cultivars

In order to facilitate the genetic improvement of NUE in GJ cultivars, we examined the haplotypes of 14 NUE genes in 41 germplasm accessions from 12 countries (Supplementary Table S16) and 71 GJ cultivars approved in China (Supplementary Table S17) using the 18 markers developed above.

We firstly analyzed the distribution of elite NUE haplotypes in 77 germplasm accessions, including the 36 accessions with reference genome (Table 2) and 41 accessions examined with developed markers (Supplementary Table S16). The elite haplotype of each NUE gene was mainly existed in XI accessions, except for that of *SBM1* and *DEP1* (Figure 3). The number of elite haplotypes carried by an XI accession ranged from 4 to 10, with an average value of 7.56. Three XI accessions, namely, IR64, R527, and YX1, carried 10 elite haplotypes. In contrast, few elite haplotypes were carried by GJ accessions, and 3 accessions carried only one elite haplotype, which were Ballila carrying allele *dep1*, and IRAT261 and Suyunuo carrying *MYB61*.

We subsequently analyzed the distribution of elite NUE haplotypes in GJ cultivars approved in the northern provinces of China. A very few elite haplotypes were carried by these cultivars, and the number of elite haplotypes carried by an accession ranged from 0 to 3. Surprisingly, the elite allele *dep1* was carried by 65 out of 71 accessions (Figure 3 and Supplementary Table S17). The HapB of *NGR5* was carried by six accessions and the HapB of *MYB61* was carried by four accessions. In addition, the HapB of *OsNPF6.1* and *NRT1.1B* was carried by an accession, respectively.

DISCUSSION

In this study, haplotype analysis of the 14 NUE genes using high-quality genomic sequences facilitated the identification of some important variations and haplotypes. For example, HapC of *ARE1* carried by XI accession 9311 contained several variations in the 5'UTR that was different from reported variations, including a 7,808-bp insertion at position –466 in this study (Supplementary Table S10) and a 6-bp insertion at position –314 in reported study (Wang Q. et al., 2018). HapD of *SBM1* was carried by 14 out of 36 accessions in this study (Supplementary Table S5), but was not identified from 1,140 rice accessions of broad genetic diversity (Xu et al., 2021). Therefore, high-quality genomic sequence was of great importance in haplotype analysis. Intragenic or functional variations that resided in the region covering from 2-kb upstream of start codon to 1-kb downstream of stop codon were selected for marker development. The markers developed in this study were ideal markers for its co-segregating with elite alleles. Haplotype analysis of the 14 NUE genes from 77 genetically

diverse rice germplasm accessions revealed that the number of elite haplotypes carried by an XI accession ranged from 4 to 10, while that carried by a GJ accession ranged from 0 to 3, which gives a good explanation of the NUE difference between the two subpopulations of Asian rice cultivars (Table 2 and Supplementary Table S16). Finally, developed intragenic markers for NUE genes and evaluated germplasm accessions in this study could be of great use in guiding the improvement of NUE in GJ cultivars in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PL performed most of the experiments and wrote the manuscript. ZL participated in marker development and materials evaluation. XL, HZ, and QW participated in materials evaluation. NL and HD provided some rice accessions and cultivars. FY designed and supervised the study. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (32101756), the Natural Science Foundation of Shandong Province (ZR2019BC105), the Key Research and Development Project of Shandong Province (2021LZGC025, 2019LZGC017, and 2019LZGC003), the Rice Industry Technology Program of Shandong (SDAIT-17-03), and the Agricultural Scientific and Technological Innovation Project of Shandong Academy of Agricultural Sciences (CXGC2021B24).

ACKNOWLEDGMENTS

We thank Prof. Guiquan Zhang from South China Agricultural University, Prof. Yuqing He and Master Qinglu Zhang from Huazhong Agricultural University, Researcher Wenfeng Chen from Guangdong Academy of Agricultural Sciences, and Associate Researcher Tianqing Zheng from Chinese Academy of Agricultural Sciences, for providing some rice germplasm accessions. We also thank Researcher Wenying Zhu and Associate Researcher Feng Chen from our institute, and Director Baizhan Yang from Seed Company of Tancheng County, for providing some rice cultivars approved in Shandong province.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.891860/full#supplementary-material>

REFERENCES

- Gao, Y., Xu, Z., Zhang, L., Li, S., Wang, S., Yang, H., et al. (2020). MYB61 is regulated by GRF4 and promotes nitrogen utilization and biomass production in rice. *Nat. Commun.* 11:5219. doi: 10.1038/s41467-020-19019-x
- Gao, Z., Wang, Y., Chen, G., Zhang, A., Yang, S., Shang, L., et al. (2019). The indica nitrate reductase gene OsNR2 allele enhances rice yield potential and nitrogen use efficiency. *Nat. Commun.* 10, 5207. doi: 10.1038/s41467-019-13110-8
- Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Sump. Ser.* 41, 95–98. doi: 10.1021/bk-1999-0734.ch008
- Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., et al. (2015). Variation in NRT1.1B contributes to nitrate-use divergence between rice subspecies. *Nat. Genet.* 47, 834–838. doi: 10.1038/ng.3337
- Huang, X., Qian, Q., Liu, Z., Sun, H., He, S., Luo, D., et al. (2009). Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* 41, 494–497. doi: 10.1038/ng.352
- International Rice Genome Sequencing Project (2005). The map-based sequence of the rice genome. *Nature* 436, 793–800. doi: 10.1038/nature03895
- Jiang, H., Li, Z., Liu, J., Shen, Z., Gao, G., Zhang, Q., et al. (2019). Development and evaluation of improved lines with broad-spectrum resistance to rice blast using nine resistance genes. *Rice (N Y)* 12:29. doi: 10.1186/s12284-019-0292-z
- Khush, G. S. (1999). Green revolution: preparing for the 21st century. *Genome* 42, 646–655. doi: 10.1139/g99-044
- Li, H., Hu, B., and Chu, C. (2017). Nitrogen use efficiency in crops: lessons from *Arabidopsis* and rice. *J. Exp. Bot.* 68, 2477–2488. doi: 10.1093/jxb/erx101
- Li, S., Tian, Y., Wu, K., Ye, Y., Yu, J., Zhang, J., et al. (2018). Modulating plant growth-metabolism coordination for sustainable agriculture. *Nature* 560, 595–600. doi: 10.1038/s41586-018-0415-5
- Liu, Y., Wang, H., Jiang, Z., Wang, W., Xu, R., Wang, Q., et al. (2021). Genomic basis of geographical adaptation to soil nitrogen in rice. *Nature* 590, 600–605. doi: 10.1038/s41586-020-03091-w
- Mao, T., Zhu, M., Ahmad, S., Ye, G., Sheng, Z., Hu, S., et al. (2021). Superior japonica rice variety YJ144 with improved rice blast resistance, yield, and quality achieved using molecular design and multiple breeding strategies. *Mol. Breed* 41:65. doi: 10.1007/s11032-021-01259-4
- Neff, M. M., Turk, E., and Kalishman, M. (2002). Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* 18, 613–615. doi: 10.1016/s0168-9525(02)02820-2
- Qin, P., Lu, H., Du, H., Wang, H., Chen, W., Chen, Z., et al. (2021). Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* 184, 3542.e–3558.e. doi: 10.1016/j.cell.2021.04.046
- Qing, D. J., Liu, K. Q., Yang, Y. Y., Gao, L. J., Huang, J., Gao, J., et al. (2018). Development of molecular marker of rice blast resistance gene *Pigm* on basis of PARMS technology. *Southwest China J. Agric. Sci.* 31, 1617–1621.
- Shen, C., Chen, K., Cui, Y., Chen, J., Mi, X., Zhu, S., et al. (2021). QTL mapping and favorable allele mining of nitrogen deficiency tolerance using an interconnected breeding population in rice. *Front. Genet.* 12:616428. doi: 10.3389/fgene.2021.616428
- Song, J. M., Xie, W. Z., Wang, S., Guo, Y. X., Koo, D. H., Kudrna, D., et al. (2021). Two gap-free reference genomes and a global view of the centromere architecture in rice. *Mol. Plant* 14, 1757–1767. doi: 10.1016/j.molp.2021.06.018
- Sun, H., Qian, Q., Wu, K., Luo, J., Wang, S., Zhang, C., et al. (2014). Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat. Genet.* 46, 652–656. doi: 10.1038/ng.2958
- Tang, W., Ye, J., Yao, X., Zhao, P., Xuan, W., Tian, Y., et al. (2019). Genome-wide associated study identifies NAC42-activated nitrate transporter conferring high nitrogen use efficiency in rice. *Nat. Commun.* 10:5279. doi: 10.1038/s41467-019-13187-1
- Wang, Q., Nian, J., Xie, X., Yu, H., Zhang, J., Bai, J., et al. (2018). Genetic variations in ARE1 mediate grain yield by modulating nitrogen utilization in rice. *Nat. Commun.* 9:735. doi: 10.1038/s41467-017-02781-w
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557, 43–49. doi: 10.1038/s41586-018-0063-9
- Wang, Y., Jiang, W., Liu, H., Zeng, Y., Du, B., Zhu, L., et al. (2017). Marker assisted pyramiding of Bph6 and Bph9 into elite restorer line 93-11 and development of functional marker for Bph9. *Rice (N Y)* 10:51. doi: 10.1186/s12284-017-0194-x
- Wu, K., Wang, S., Song, W., Zhang, J., Wang, Y., Liu, Q., et al. (2020). Enhanced sustainable green revolution yield via nitrogen-responsive chromatin modulation in rice. *Science* 367:eaaz2046. doi: 10.1126/science.aaz2046
- Xu, J., Shang, L., Wang, J., Chen, M., Fu, X., He, H., et al. (2021). The SEEDLING BIOMASS 1 allele from indica rice enhances yield performance under low-nitrogen environments. *Plant Biotechnol. J.* 19, 1681–1683. doi: 10.1111/pbi.13642
- Yu, J., Xuan, W., Tian, Y., Fan, L., Sun, J., Tang, W., et al. (2021). Enhanced OsNLP4-OsNiR cascade confers nitrogen use efficiency by promoting tiller number in rice. *Plant Biotechnol. J.* 19, 167–176. doi: 10.1111/pbi.13450
- Zhang, H., Wang, Y., Deng, C., Zhao, S., Zhang, P., Feng, J., et al. (2022). High-quality genome assembly of Huazhan and Tianfeng, the parents of an elite rice hybrid Tian-you-hua-zhan. *Sci. China Life Sci.* 65, 398–411. doi: 10.1007/s11427-020-1940-9
- Zhang, S., Zhu, L., Shen, C., Ji, Z., Zhang, H., Zhang, T., et al. (2021). Natural allelic variation in a modulator of auxin homeostasis improves grain yield and nitrogen use efficiency in rice. *Plant Cell* 33, 566–580. doi: 10.1093/plcell/koaa037
- Zhang, Y., Tan, L., Zhu, Z., Yuan, L., Xie, D., and Sun, C. (2015). TOND1 confers tolerance to nitrogen deficiency in rice. *Plant J.* 81, 367–376. doi: 10.1111/tpj.12736
- Zhou, Y., Tao, Y., Tang, D., Wang, J., Zhong, J., Wang, Y., et al. (2017). Identification of QTL associated with nitrogen uptake and nitrogen use efficiency using high throughput genotyped CSSLs in rice (*Oryza sativa* L.). *Front. Plant Sci.* 8:1166. doi: 10.3389/fpls.2017.01166

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Li, Liu, Zhang, Wang, Li, Ding and Yao. 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(s) 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.



Small RNA and Degradome Deep Sequencing Reveals the Roles of microRNAs in Peanut (*Arachis hypogaea* L.) Cold Response

Xin Zhang^{1,2}, Chao Ren¹, Yunyun Xue¹, Yuexia Tian¹, Huiqi Zhang¹, Na Li¹, Cong Sheng^{3,4}, Huifang Jiang^{5*} and Dongmei Bai^{1*}

OPEN ACCESS

Edited by:

Chuanzhi Zhao,
Shandong Academy of Agricultural
Sciences, China

Reviewed by:

Bingyan Huang,
Henan Academy of Agricultural
Sciences, Zhengzhou, China
Xiaoyuan Chi,
Shandong Peanut Research Institute
(CAAS), China

*Correspondence:

Dongmei Bai
baidm1221@163.com
Huifang Jiang
peanutlab@oilcrops.cn

Specialty section:

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

Received: 14 April 2022

Accepted: 06 May 2022

Published: 02 June 2022

Citation:

Zhang X, Ren C, Xue Y, Tian Y,
Zhang H, Li N, Sheng C, Jiang H and
Bai D (2022) Small RNA and
Degradome Deep Sequencing
Reveals the Roles of microRNAs in
Peanut (*Arachis hypogaea* L.) Cold
Response.
Front. Plant Sci. 13:920195.
doi: 10.3389/fpls.2022.920195

¹Institute of Industrial Crops, Shanxi Agricultural University, Taiyuan, China, ²State Key Laboratory of Sustainable Dryland Agriculture, Shanxi Agricultural University, Taiyuan, China, ³Department of Plant Pathology, College of Plant Protection, Nanjing Agricultural University, Nanjing, China, ⁴Laboratory of Bio-interactions and Crop Health, Nanjing Agricultural University, Nanjing, China, ⁵Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, Wuhan, China

Cold stress is a major environmental factor that affects plant growth and development, as well as fruit postharvest life and quality. MicroRNAs (miRNAs) are a class of non-coding small RNAs that play crucial roles in various abiotic stresses. Peanuts (*Arachis hypogaea* L.), one of the most important grain legumes and source of edible oils and proteins, are cultivated in the semi-arid tropical and subtropical regions of the world. To date, there has been no report on the role of miRNAs in the response to cold stress in cultivated peanuts. In this study, we profiled cold-responsive miRNAs in peanuts using deep sequencing in cold-sensitive (WQL20) alongside a tolerant variety (WQL30). A total of 407 known miRNAs and 143 novel peanut-specific miRNAs were identified. The expression of selected known and novel miRNAs was validated by northern blotting and six known cold-responsive miRNAs were revealed. Degradome sequencing identified six cold-responsive miRNAs that regulate 12 target genes. The correlative expression patterns of several miRNAs and their target genes were further validated using qRT-PCR. Our data showed that miR160-ARF, miR482-WDR1, miR2118-DR, miR396-GRF, miR162-DCL, miR1511-SRF, and miR1511-SPIRAL1 modules may mediate cold stress responses. Transient expression analysis in *Nicotiana benthamiana* found that miR160, miR482, and miR2118 may play positive roles, and miR396, miR162, and miR1511 play negative roles in the regulation of peanut cold tolerance. Our results provide a foundation for understanding miRNA-dependent cold stress response in peanuts. The characterized correlations between miRNAs and their response to cold stress could serve as markers in breeding programs or tools for improving cold tolerance of peanuts.

Keywords: peanut, microRNA, cold stress, cold response, cold tolerance

INTRODUCTION

Cultivated peanuts (*Arachis hypogaea* L.) are vital oil plants and cash crops that globally play an essential role in consuming edible vegetable oil and leisure food. However, low temperatures limit the scope of crops to a large extent and cause yield reduction, quality decline, and occasionally no harvest in severe cases (Jeon and Kim, 2013; Barrero-Gil and Salinas, 2018). The annual loss of crops caused by low-temperature damage worldwide is as high as hundreds of billions of dollars (Xiao and Song, 2011; Abba et al., 2019). Therefore, the improvement of cold tolerance has essential application value for crops and other plants and is an urgent research topic.

As a thermophilic crop, peanuts require relatively high temperatures throughout their development. Cold injury causes different degrees of harm to the peanut at seedling emergence, flowering, maturity, and other key growth stages. Among them, the impact of the seedling stage is the most common, which could delay peanut germination and seedling growth, resulting in uneven seedling emergence, weak seedling potential, prolonged seedling emergence time, lack of seedlings, and ridge cutting. Research on cold stress in plants began in the early 1830s and today has a history of more than 190 years. Breeders have been trying to develop new varieties to resolve the problem of peanut chilling damage and have made progress. A few cold-tolerant and early maturing varieties have been cultivated (Upadhyaya et al., 2003; Zhang et al., 2019). However, the mechanism of cold tolerance is a complex quantitative trait that always appears in combination or continuously and is not controlled by a single regulatory pathway or gene, making traditional cold tolerance breeding methods challenging with longer breeding cycles (Park et al., 2015; Shi et al., 2018; Zhang et al., 2019). Therefore, understanding peanut cold defense mechanisms is necessary to accelerate the cultivation of peanut varieties with high cold tolerance. Under low non-freezing temperature conditions, plants reprogram their gene expression through transcriptional, post-transcriptional, and post-translational mechanisms and undergo physiological and biochemical adjustments to improve their tolerance to low temperatures. Many studies have shown that the molecular mechanisms of cold defense in many plants are evident. The plant's cold defense molecular mechanism is a highly complex network composed of numerous positive and negative regulatory factors, which can form a center that integrates a variety of internal and external signals by binding to the C-repeat binding transcription factor (CBF)/dehydrate responsive element binding factor (DREB) promoter (Barrero-Gil and Salinas, 2018). The core mechanism is that cold stress activates the plant's ICE-CBF-COR regulatory pathway. Under cold stress, the transcription factor CBF responds to cold signals with the participation of the transcription activator ICE (inducer of CBF expression), binds to the cis-acting element CRT (C-repeat)/DER (dehydration) of the cold-induced genes COR (cold-regulated) responsive element sequence, initiates the expression of a series of cold-induced genes and proteins, and then changes the physiology, metabolism, and growth of plants (Gilmour et al., 1998;

Wang et al., 2017; Barrero-Gil and Salinas, 2018; Shi et al., 2018; Guo et al., 2019). The ICE-CBF-COR transcriptional cascade has been established as the main regulatory response toolkit for cold signaling and freezing tolerance. Peanuts are sensitive to low temperatures, but they have evolved a unique low-temperature survival mechanism (Zhu, 2016). Studies have reported that overexpression of *AtDREB2A* in peanuts can improve their tolerance to low temperature, drought, and salt stress (Pruthvi et al., 2014). *AhPNDREB1* (a homologous gene of *AtCBF1*), which is intron-less and constitutively expressed, was isolated from the cultivar Luhua 14 (Zhang et al., 2009). The AP2 domain in *AhPNDREB1* is specific to the CRT/DER element combining ability, and the gene can be vigorously and rapidly induced to be expressed under cold and drought stress. Chen et al. (2012) cloned peanut *AhCBF2* and *AhCBF15*, analyzed their amino acid sequences, and found that they both contained the unique ERF/AP2 domain of CBF/DREB family genes. However, there are few reports on the mechanism of peanut cold tolerance, which requires further exploration.

Over the past few years, many studies have shown that small RNA (sRNA) play a pivotal role in regulating cold stress in various plants. sRNA, as a type-21-24 nucleotide long non-coding RNA molecule, is widespread in eukaryotes and has two main forms: microRNA (miRNA) and small interfering RNA (siRNA) (Jin, 2008). It can form an RNA-induced silencing complex by combining with the ribozyme complex containing the AGO protein to cleave target mRNA or inhibit its translation and is widely involved in the cold stress responses of various plants, such as *Arabidopsis* (Sunkar and Zhu, 2004; Zhou et al., 2008; Yan et al. 2016), poplar (Lu et al., 2008), wheat (Tang et al., 2012), rice (Yang et al., 2013), soybean (Zhang et al., 2014), tomato (Shi et al., 2019), and wild grape (Wang et al., 2019). By targeting stress-related transcription factors, such as MYB, WRKY, and bHLH (Zhang et al., 2014; Shi et al., 2019; Wang et al., 2019), genes related to the reactive oxygen scavenging system (Yang et al., 2013) or ABA signaling pathway genes (Yan et al., 2016), these miRNAs may regulate the cold tolerance of plants. Studies have shown that sRNA-mediated gene silencing, histone modification, phosphorylation, and ubiquitination can regulate CBF expression at post-transcriptional and post-translational levels (Barrero-Gil and Salinas, 2018; Lantzouni et al., 2020). Thus, we believe that exploring the peanut cold defense regulatory network from small RNA-mediated gene silencing will be helpful in revealing the peanut response and regulatory mechanisms of cold stress.

To understand the role of miRNAs in peanut cold tolerance, we used two representative peanut lines in this study: WQL20 (waiting for quiz line 20) and WQL30 were eighth-generation recombinant inbred lines (RILs) from a cross between cultivars of Yuanza 9102 (female) and Xuzhou 68-4 (male). The main difference between the two RILs is their ability to withstand low temperatures: WQL30 is resistant to low temperatures, whereas WQL20 is susceptible to low temperatures. These two peanut lines were investigated to understand peanut cold tolerance by analyzing plant miRNA expression patterns and their target genes.

MATERIALS AND METHODS

Plant Materials

The RIL population from a cross between cultivars of Yuanza 9102 (female) and Xuzhou 68-4 (male) includes 188 F₈ lines. Yuanza 9102, the female parent of the RIL population, belonging to *A. hypogaea* subsp. *fastigiata* var. *vulgaris*, was derived from interspecific hybridization between cultivated cultivar Baisha 1,016 and a diploid wild species *A. diogoi*. Xuzhou 68-4, the male parent, which belongs to *A. hypogaea* subsp. *hypogaea* var. *hypogaea*. Yuanza 9102 is a popular cultivar in China that is resistant to cold stress, while Xuzhou 68-4 is susceptible to cold stress.

We conducted hydroponic culture of the two lines in Hoagland solution in a growth chamber, and then culture them in illumination incubator in the laboratory. The temperature was 27°C and the light-dark cycle was 14 h of light and 10 h of darkness. For low-temperature treatment, 6 four-leaf stage seedlings for each line were kept at 6°C, 70% RH, and a 14-h/10-h light/dark photoperiod. Lines WQL30 and WQL20 showed apparent difference resistance to cold. After 4 h of cold treatment (HCT), WQL20 began to wilt, while WQL30 did not change significantly. After 8 h of cold treatment (HCT), WQL20 wilted significantly, but WQL30 still had no noticeable change. After 12 h of cold treatment (HCT), WQL20 wilted and fell, while WQL30 showed slight wilting symptoms (**Figure 1**). Samples of four HCT (0 h, 4 h, 8 h, and 12 h) were collected from whole seedlings for miRNAome and degradome analysis.

sRNA Library Construction and DNA Sequencing

RNA degradation and contamination was monitored on 1% agarose gels. RNA purity was checked using the NanoPhotometer[®] spectrophotometer (IMPLEN, CA, United States). RNA concentration was measured using Qubit[®] RNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, CA, United States). A total amount of 3 µg total RNA per sample was used as input material for the small RNA library. Sequencing libraries were generated using NEBNext[®] Multiplex Small RNA Library Prep Set for Illumina[®] (NEB, United States) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, Total RNA was ligated to the RNA 3' and RNA 5' adapters and reverse transcription, followed by PCR, was performed to make cDNA constructs of the small RNAs. PCR products were purified on a 8% polyacrylamide gel (100 V, 80 min). DNA fragments corresponding to 140~160 bp (the length of small non-coding RNA plus the 3' and 5' adaptors) were recovered and dissolved in 8 µl elution buffer. At last, library quality was assessed on the Agilent Bioanalyzer 2,100 system using DNA High Sensitivity Chips. We then performed single-end sequencing (50 bp) on an Illumina HiSeq2500 at the LC-BIO (Novogene Experimental Department, Beijing, China) following the vendor's recommended protocol.

Degradome Library Construction and Sequencing

To identify the potential targets, equal amounts of RNAs from WQL20 (0 HCT, 4 HCT, 8 HCT, and 12 HCT) and WQL30 samples (0 HCT, 4 HCT, 8 HCT, and 12 HCT) were mixed together for degradome library construction and deep sequencing. The two libraries named as DS1 and DS2, respectively. Through preprocessing, clean tags are generated. Then, clean tags were classified by alignment with GenBank, Rfam database, and miRNA database. Next, the reads were mapped to the reference genome.¹ The sense strand of peanut cDNA was used to predict miRNA cleavage sites using CleaveL and pipeline (Addo-Quaye et al., 2009). Based on the signature number and abundance of cleaved position at each occupied transcript, the cleaved transcripts could be categorized into five categories (0, 1, 2, 3, and 4).

Northern Blotting Analysis

RNA blot analyses for miRNAs from total extracts were performed as described previously (Katiyar-Agarwal and Jin, 2007). Total RNA was extracted using TRIzol reagent (Takara, Japan) following the manufacturer's instructions. RNA was resolved on a 14% denaturing 8M urea-PAGE gel and then transferred and chemically crosslinked onto a Hybond N⁺ membrane (GE Healthcare Life Sciences) using N-(3-Dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride. miRNA probes were end-labeled with [³²P] ATP by T4 polynucleotide kinase (New England Biolabs). For internal control experiments, blots were stripped for 20 min and 4 times at 80°C in stripping buffer (0.1 X SSC/0.5% SDS). After detecting no signal, the stripped blots were rehybridized with r-³²P ATP-labeled U6 gene fragments to confirm loading amounts. All signals were normalized to the U6 signals obtained from each blot. Expression levels were quantified using Image J as instructed.

qRT-PCR

The RNA used for qRT-PCR was from the same set RNA for northern blotting analysis. The RNA was then reverse transcribed into cDNA using the PrimeScript RT reagent Kit (Takara, Japan). The qRT-PCR was performed in a 15 µl reaction mixture consisting of 1.5 µl 1×SYBR Green (Invitrogen, United States), 1.5 µl PCR buffer, 0.3 µl 10 mM dNTPs (Takara, Japan), 0.3 µl Taq, 0.3 µl ROX DYE2 (Vazyme, China), 1.5 µl 2 mM each primer and 2 µl appropriate diluted cDNA. The conditions for qRT-PCR were as follow: 94°C for 3 min, then 40 cycles at 94°C for 30 s and 58°C for 30 s, followed by 72°C for 35 s for PCR amplification. Transcript levels of each gene were measured by the Applied Biosystems 7500 (Applied Biosystems, United States) according to the manufacturer's instructions. 18S rRNA was used as a quantitative control in the qRT-PCR analysis. Primers used in this study are listed in **Supplementary Table S4**.

¹<http://www.peanutbase.org>

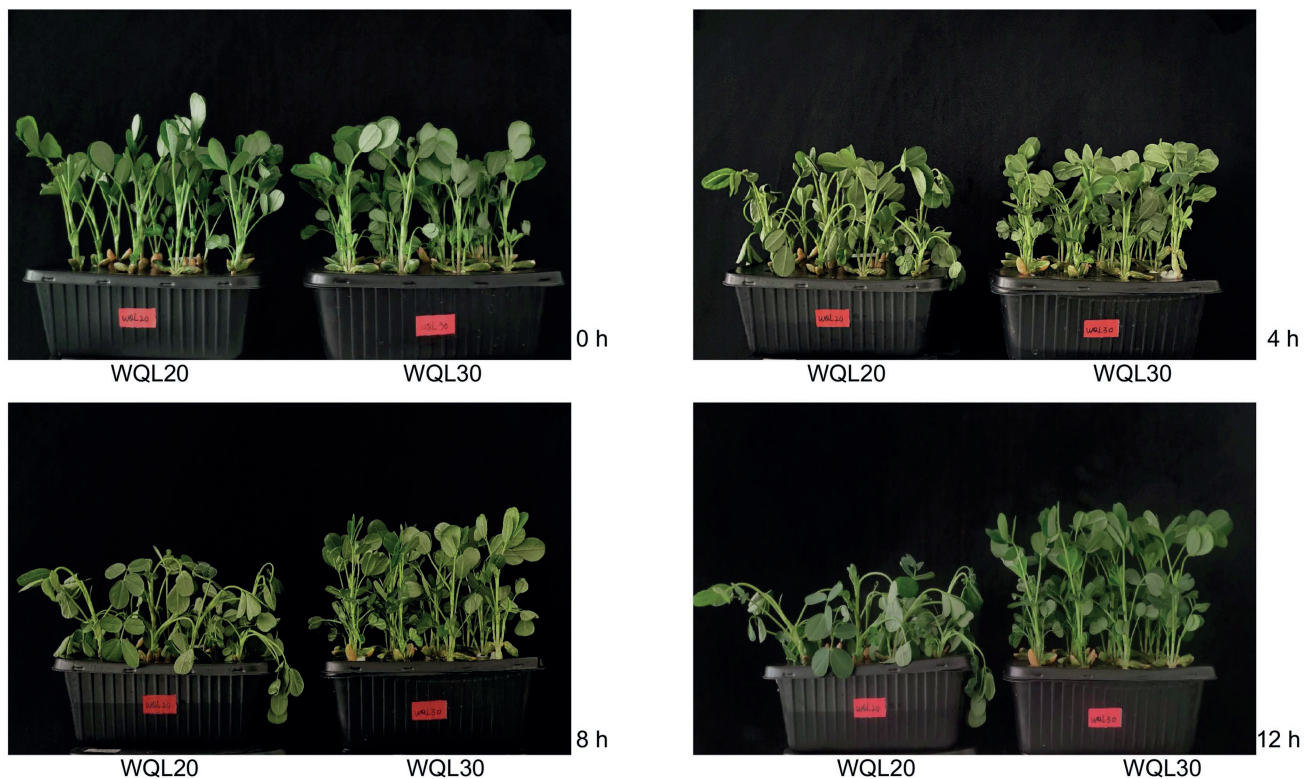


FIGURE 1 | The phenotype of peanut exposed to low temperature. For low-temperature treatment, four-leaf stage seedlings of WQL30 and WQL20 were kept in a growth chamber at 6°C, 70% RH, and a 14-h/10-h light/dark photoperiod. With the increase of treatment time, WQL20 began to wilt, while WQL30 did not change significantly after 4 h of cold treatment. After 8 h of cold treatment, WQL20 wilted significantly, but WQL30 still had no noticeable change. After 12 h of cold treatment, WQL20 wilted and fell, while WQL30 showed slight wilting symptoms as shown in this figure. The pictures were taken at 0 HCT, 4 HCT, 8 HCT, and 12 HCT and assembled in Adobe Photoshop CS6.

Transient Expression Analysis of miRNA in *Nicotiana benthamiana*

The precursor of miRNA, *MIRNA*, cloned from peanut was sub-cloned into the overexpression vector pCambia1300 destination vector using LR clonase II (Invitrogen). Transient co-expression assays in *N. benthamiana* were performed by infiltrating 4-week-old *N. benthamiana* plants with *Agrobacterium* GV3101 [OD₆₀₀ (optical density at 600 nm)=1.0] harboring constructs containing the *MIRNA* (pCambia1300). The *Agrobacterium*-mediated transformation was performed according to a previously described method (Niu et al., 2018; Zhang et al., 2018). After cultivation for 48 h under 16 h light/8 h dark at 23°C conditions, the injected tobacco plants were treated with 4°C stress for 24 h in an incubator (Dongqi, Ningbo, China). Then the injected tobacco leaves were collected for RNA extraction.

RESULTS

High-Throughput Sequencing and Annotation of Peanut sRNAs

To identify endogenous cold stress-responsive miRNAs that are potentially involved in cold response, we treated the two

Lines (WQL20 and WQL30) with low-temperature 6°C and sampled tissue from treated plants at four different time points after treatment [0, 4, 8, and 12 h after cold treatment (HCT)]. Total RNA from each sample was used for sRNA library construction. As shown in Table 1, 9,576,091, 8,391,121, 9,343,483, and 7,954,268 total reads were generated from each library representing the different time points in line WQL20, respectively. 9,376,386, 11,205,178, 10,219,016, and 8,626,070 total reads were in line WQL30, respectively. By mapping to the peanut genome (Cultivated peanut, *A. hypogaea* cv. Tifrunner),² 9,349,005, 8,234,001, 9,129,232 and 7,784,811 peanut sRNA reads were obtained, corresponding to 2,093,285, 2,057,134, 2,270,139 and 1,634,313 unique reads in WQL20, respectively. 9,242,332, 11,025,444, 9,875,961 and 8,492,492 peanut sRNA reads were obtained in WQL30, corresponding to 1,833,592, 2,107,138, 2,030,541 and 1,998,859 unique reads, respectively. These reads were further searched against the Rfam database³ to remove known sRNA, such as ribosomal RNA (rRNA), transfer RNA (tRNA), and small nucleolar RNAs (snoRNAs), the Repbase databases⁴ to remove repeats. In the end, 5,676,589,

²<http://www.peanutbase.org>

³<http://xfam.org/>

⁴<http://www.girinst.org/repbase/update/index.html>

TABLE 1 | Distribution of sRNAs among different categories in each library.

Types	Normalized Reads							
	WQL20				WQL30			
	0h	4h	8h	12h	0h	4h	8h	12h
Total sRNA	9,576,091	8,391,121	9,343,483	7,954,268	9,376,386	11,205,178	10,219,016	8,626,070
Mapped sRNA	9,349,005	8,234,001	9,129,232	7,784,811	9,242,332	11,025,444	9,875,961	8,492,492
Unique reads	2,093,285	2,057,134	2,270,139	1,634,313	1,833,592	2,107,138	2,030,541	1,998,859
Rfam								
rRNA	3,024,554	2,104,932	2,902,873	2,898,576	3,674,164	4,202,996	3,637,792	2,867,982
tRNA	144,762	152,863	124,882	215,951	154,697	292,961	187,397	146,373
snRNA	918	844	1,586	873	861	1,584	847	829
snoRNA	13,669	15,491	38,404	16,492	16,673	24,703	13,390	13,998
Repeat	488,513	632,601	626,393	603,585	431,565	520,481	485,189	443,443
Peanut small RNA	5,676,589	5,327,270	5,435,094	4,049,334	4,964,372	5,982,719	5,551,346	5,019,867
Conserved_miRNA (Total)	1,089,490	942,979	359,409	555,137	1,004,326	1,207,299	1,090,170	881,849
Conserved_miRNA (Unique)	1,328	1,260	1,126	1,184	1,295	1,414	1,339	1,233
Novel_miRNA (Total)	58,424	70,267	34,750	42,150	65,449	103,589	84,218	73,171
Novel_miRNA (Unique)	1,243	1,439	1,398	1,166	1,370	1,546	1,536	1,386
TAS (Total)	3,094	6,961	4,133	2,944	3,508	5,193	5,104	4,408
TAS (Unique)	300	366	298	267	292	333	297	304
NAT (Total)	654,060	757,250	904,777	694,258	687,859	852,495	774,593	663,985
NAT (Unique)	157,239	181,283	199,877	147,463	153,205	180,847	169,278	161,026
exon:+	206,065	230,953	266,696	164,444	183,927	223,993	201,463	170,121
exon:-	126,392	148,788	134,160	99,348	120,449	169,910	139,523	118,291
intron:+	172,157	214,436	249,986	156,912	171,819	209,054	203,835	188,352
intron:-	103,208	127,995	154,537	96,073	109,887	123,634	124,446	120,489
other	3,263,699	2,827,641	3,326,646	2,238,068	2,617,148	3,087,552	2,927,994	2,799,201

A peanut sRNA deep sequencing summary is presented. The abundance in different libraries was normalized to that in the library at 0h, and sRNAs were calculated as reads per million.

5,327,270, 5,435,094 and 4,049,334 reads were retrieved from each library of WQL20, and 4,964,372, 5,982,719, 5,551,346 and 5,019,867 reads were retrieved from each library of WQL30. The reads were considered as sRNA originated.

The lengths of the unique, valid reads ranged from 18 to 30 nucleotides (nt), and the 21–24 nt sequences were predominant in each library, with the 24 (nt) sequences being the most common (**Supplementary Figure S1**).

Identification of Known, Conserved, and Novel miRNAs

To identify known, conserved, and novel miRNAs in peanut, all the total peanut sRNAs were classified into three categories according to their origination and structural features: (1) conserved miRNAs with a hairpin structure of their precursor sequences were aligned to plant miRNAs in miRBase⁵; (2) novel peanut miRNAs have a hairpin structure of their precursor sequences but are not listed in the microRNA database; and (3) peanut siRNAs that do not have a hairpin structure of their precursor sequences. According to this criteria, we retrieved 1,328, 1,260, 1,126, and 1,184 unique conserved peanut miRNAs in each library of WQL20 and 1,295, 1,414, 1,339, and 1,233 unique conserved peanut miRNAs in each library of WQL30, respectively (**Table 1**; **Supplementary Table S1**). In addition,

we obtained 1,243, 1,439, 1,398, and 1,166 novel peanut miRNAs in each library of WQL20 and 1,370, 1,546, 1,536, and 1,386 novel peanut miRNAs in each library of WQL30, respectively.

Comparison of Differentially Expressed miRNAs Between the Two Peanut RILs

To identify the differentially expressed miRNAs, we compared the frequencies of occurrence of differentially expressed miRNAs at the 0 HCT, 4 HCT, 8 HCT, and 12 HCT stages between the two lines after the miRNA reads were normalized to transcripts per million (TPM). We identified 56, 71, and 157 differentially expressed miRNAs in the WQL20_4h vs. WQL20_0h, WQL20_8h vs. WQL20_0h, and WQL20_12h vs. WQL20_0h comparisons, and 23, 23, and 23 differentially expressed miRNAs in the WQL30_4h vs. WQL30_0h, WQL30_8h vs. WQL30_0h and WQL30_12h vs. WQL30_0h comparisons, respectively (**Figure 2A**). Among these, 39, 96, and 36 miRNAs were upregulated expression, and 17, 61, and 35 miRNAs were downregulated at 4 HDT, 8 HDT, and 12 HDT compared to 0 HDT in RIL WQL20, respectively. 9, 12, and 12 miRNAs were upregulated expression, and 14, 11, and 11 miRNAs downregulated expression in RIL WQL30, respectively (**Figure 2B**). To identify the miRNAs responsive to cold stress, the miRNAs with opposite changes between the two peanut RILs both abundantly and significantly differentially expressed were identified by employing the following criteria: (1) total reads $\geq 3,000$; (2) [treated/mock] ≥ 2 or [treated/mock] ≤ 0.5 in

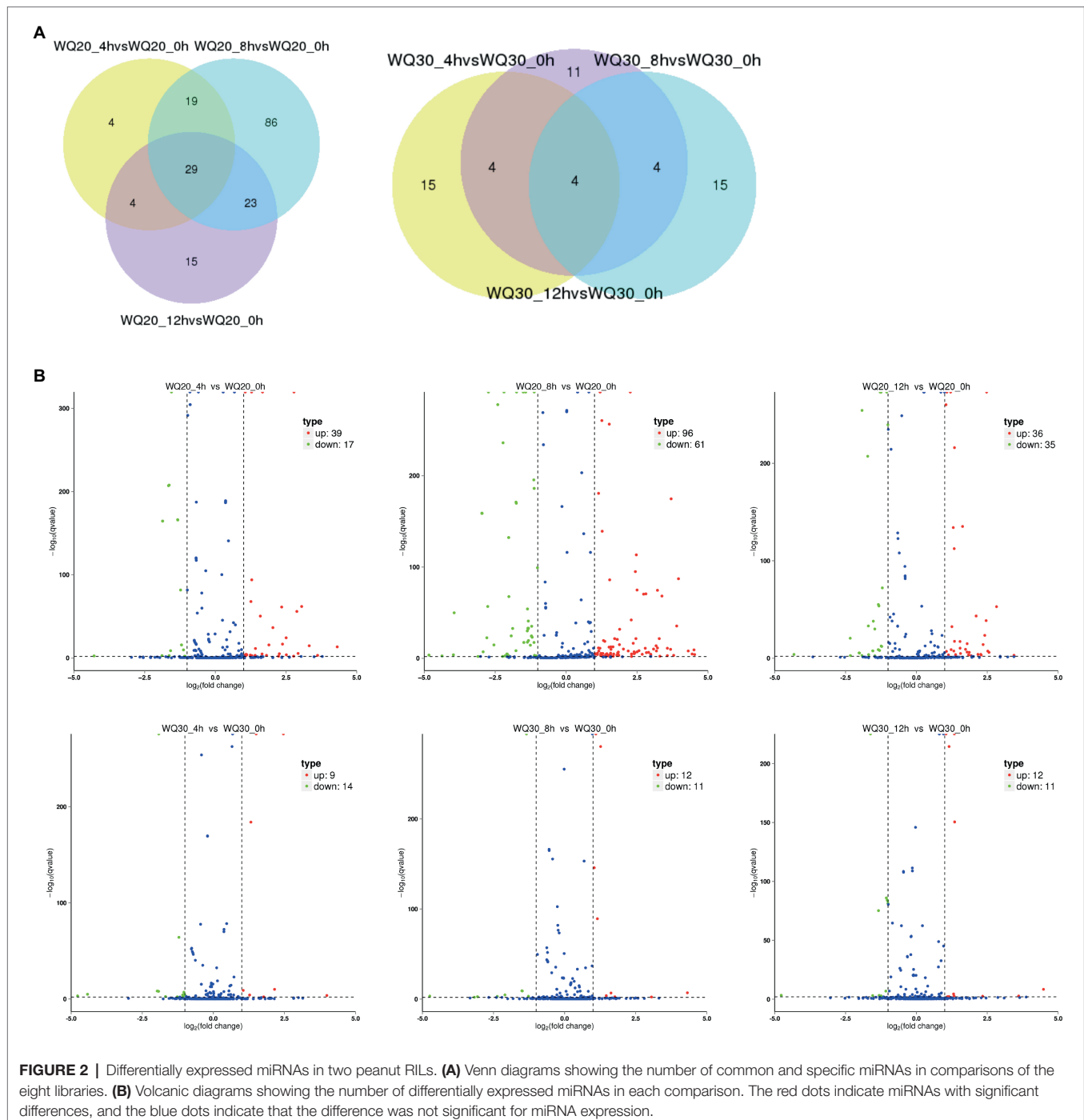
⁵<http://www.mirbase.org/>

at least one stage. In the end, we found 12 significantly differentially expressed conserved peanut miRNAs (DE-miRNAs; Table 2).

Validation of Cold-Responsive miRNAs by Northern Blotting

By using reverse-cDNA fragments as probes, we validated our bioinformatics prediction by northern blotting assay (there is only

one nucleotide difference between ath-miR396a-5p and gma-miR396h, between ath-miR396a-5p and vvi-miR396a, between pvu-miR2118 and gma-miR2118a-3p; therefore, detection using probes reverse complementary to the aforementioned sRNAs represents the expression of the family). The eight northern blottings that were used to detect the candidate miRNAs. ahy-miR2118 and ahy-miR160 showed discernible downregulation in line WQL20 and upregulation in line WQL30 (Figures 3A,B). ahy-miR482 showed discernible downregulation in line WQL20 and weak



upregulation in line WQL30 (**Figure 3B**). ahy-miR1511 showed discernible upregulation in line WQL20 and weak downregulation in line WQL30 (**Figure 3B**). ahy-miR396 and ahy-miR162 showed discernible upregulation in line WQL20 and downregulation in line WQL30 (**Figure 3C**). Surprisingly, ahy-miR167 showed strong expression with no noticeable expression variation, which is different to the bioinformatic prediction (**Figure 3B**). ahy-miR156 expressed under the detectable level (**Figure 3A**).

Identification of Target Genes of miRNAs by Degradome Analysis

To further understand the biological functions of cold-responsive miRNAs, the miRNA target genes were identified by Degradome Analysis. In this study, we constructed two degradome libraries, DS1 (including WQL20_0h, WQL20_4h, WQL20_8h, and WQL20_12h) and DS2 (including WQL30_0h, WQL30_4h, WQL30_8h, and WQL30_12h). In total, we obtained 34,299,289 and 43,309,050 raw reads from DS1 and DS2, respectively. After removing the reads without the 3' adaptor sequence, we obtained 10,139,640 and 9,850,766 unique raw reads from the DS1 and DS2 libraries, respectively. The unique reads were aligned to the peanut genome database, and 25,396,751 (74.04%) and 30,399,096 (70.19%) reads were mapped to the genome, respectively. The mapped reads represented 7,005,430 (69.09%) and 6,538,802 (66.38%) annotated peanut genes in the DS1 and DS2 libraries, respectively (**Supplementary Table S2**). We identified cleaved targets for miRNAs based on a method in the Cleaveland pipeline (Addo-Quaye et al., 2009), in which a host gene with an alignment score of 7 or less was considered to be a potential target. According to the signature number and abundance of putative cleaved position at each occupied transcript, these cleaved transcripts could be categorized into five classes (0, 1, 2, 3, and 4). In total, 7,561 and 7,296 targets were identified from the DS1 and DS2 libraries, respectively (**Supplementary Table S3**). For these targets, 291, 70, 2,421, 991, and 3,788 were classified to categories 0, 1, 2, 3, and 4 in the DS1 library, and 365, 41, 2,358, 1,188, and 3,344 belonged to a category <4 in the DS2 library (**Supplementary Table S3**). There were 11,615 differentially expressed target genes between the DS1 and DS2 libraries. 6,317 genes were upregulated expression for these targets, and 5,298 genes were downregulated expression (**Supplementary Table S3**).

For the cold-responsive miRNAs, we found that ahy-miR160 targets auxin response factor genes *AhARF10* (Arahy.4D5DQA) and *AhARF17* (Arahy.35XBQI) (**Figures 4A,D**); ahy-miR162 targets DEAD-box ATP-dependent RNA helicase genes *AhDCL6* (Arahy.GMG8MT) and *AhDCL16* (Arahy.NHK71Z) (**Figures 4B,E**); ahy-miR396 targets the growth-regulating factor 1 genes *AhGRF1* (Arahy.858S2J) and *AhGRF4* (Arahy.KV1X6M) (**Figures 4C,F**); ahy-miR482 targets WD repeat-containing protein 26-like genes *AhWDR1* (Arahy.C265D9) and *AhWDR2* (Arahy.ER1LI4) (**Figures 4G,J**); ahy-miR1511 targets protein kinase superfamily protein gene *AhSRF* (Arahy.XK3KPA) and SPIRAL1-like 1 gene *AhSPIL1* (Arahy.NV23DS) (**Figures 4H,K**); ahy-miR2118 targets disease resistance protein genes *AhDR1* (Arahy.1ALG22) and *AhDR2* (Arahy.GT0Q5X) (**Figures 4I,L**).

We further examined the *in vivo* expression of each predicted target gene upon low-temperature treatment. The RNA used for qRT-PCR was from the same set RNA for northern blot analyses. Both of the transcriptions of *ARF10* and *ARF17* (targeted by ahy-miR160) were significantly upregulated expression in line WQL20, but significantly downregulated expression in line WQL30, respectively. Same as *AhGRF1* and *AhGRF4* (targeted by ahy-miR396), *AhDCL6* and *AhDCL16* (targeted by ahy-miR162) showed similar expression profiles between the two RIL lines. In line WQL20, they were significantly downregulated expression, but significantly upregulated expression in line WQL30, respectively (**Figure 5A**). *AhDR1* and *AhDR2*, target genes of ahy-miR2118, showed sustained increase with the low-temperature treatment in line WQL20, while sustained decrease in line WQL30, respectively. The transcriptions of *AhSRF* and *AhSPIL1* showed a noticeable reduction since 4 h and reached less than 40% of the original level at 8 h in line WQL20, but significantly increase in WQL30 with low-temperature treatment, which corresponds to the overall expression tendency of ahy-miR1511. The expressions of *AhWDR1* and *AhWDR2* are cognately inverse to the expression pattern of ahy-miR482, further suggesting a regulatory relationship between these two parties (**Figure 5B**). In summary, our research indicates that these genes are authentic target genes of their cognate miRNAs. These miRNAs may be involved in the cold-responsive process in peanut by negatively regulating their target genes.

Functional Analysis of the Six Cold-Responsive miRNAs in Tobacco

To further analyze the function of cold-responsive miRNAs, we transiently overexpressed the six cold-responsive miRNAs in tobacco leaves using *Agrobacterium* expressing pCAMBIA1300-miRNAs (**Figure 6A**). After 24 h cold treatment, the tobacco plants showed visual phenotypes. Compared with the control, the leaves overexpressing ahy-miR160, ahy-miR482, and ahy-miR2118 exhibited better growth status, but the leaves overexpressing ahy-miR396, ahy-miR162, and ahy-miR1511 showed poor growth status (**Figures 6B,C**). These results suggest that tobacco leaves overexpressing ahy-miR160, ahy-miR482, and ahy-miR2118 have enhanced cold tolerance. On the contrary, tobacco leaves overexpressing ahy-miR396, ahy-miR162, and ahy-miR1511 have weakened cold tolerance.

DISCUSSION

The Cold-Responsive miRNAs in Peanut

Cold stress is the main limiting factor for the distribution, yield, and quality of crops (Zhang et al., 2019). The cold signaling pathway has been well studied (Jeon and Kim, 2013; Barrero-Gil and Salinas, 2018; Shi et al., 2018), and many studies have shown that miRNAs play critical roles in cold stress in plants. For example, osa-miR319 can silence *OsPCF5* and *OsPCF8*, and positively regulate cold tolerance in rice (Yang et al., 2013); osa-miR156k targets the cold stress-related

TABLE 2 | List of some candidate miRNAs involved in low-temperature response in each library.

miRNA Identifier	Normalized reads (reads/mgs)						Fold change					
	WQL20			WQL30			WQL20			WQL30		
	0h	4h	8h	12h	0h	4h	4h/0h	8h/0h	12h/0h	4h/0h	8h/0h	12h/0h
ahy-miR396	12078.55	17044.83	18074.27	27523.55	16443.92	14907.85	1.41	1.50	2.28	0.91	0.89	0.84
gma-miR396a-5p	12091.90	17038.83	18086.71	27498.87	16440.42	14907.85	1.41	1.50	2.27	0.91	0.89	0.84
gma-miR396b	12058.79	17032.24	18064.93	27494.12	16426.41	14890.62	1.41	1.50	2.28	0.91	0.89	0.84
zma-miR396g-3p	4968.34	6500.29	7497.62	10354.21	3998.15	2943.47	1.31	1.51	2.08	0.74	0.74	0.66
wi-miR396a	7017.83	16039.02	18027.59	18560.48	17306.84	16602.84	2.29	2.57	2.64	0.96	1.11	0.93
ahy-miR167	2425.43	8460.94	17159.43	15457.25	10430.93	10406.25	3.49	7.07	6.37	1.00	0.86	0.87
ahy-miR1511	10571.60	3956.68	7737.22	6511.26	8098.48	10688.83	0.37	0.73	0.62	1.32	1.15	1.07
ahy-miR482	1911.19	576.38	692.35	950.53	1144.33	6018.87	0.30	0.36	0.50	5.26	2.03	2.41
ahy-miR2118	1911.19	576.38	692.35	950.53	1143.75	6018.87	0.30	0.36	0.50	5.26	2.03	2.42
gma-miR2118a-3p	978.82	758.11	356.29	997.06	743.82	1090.94	0.77	0.36	1.02	1.47	1.79	1.73
mdm-miR156ad	656.29	915.25	1754.99	1099.61	629.97	554.33	1.39	2.67	1.68	0.88	0.76	0.65
bdi-miR162	300.11	267.50	231.82	147.18	280.24	338.21	0.89	0.77	0.49	1.21	1.35	1.21
ahy-miR160												

Variations in expression were analyzed. Peanut miRNAs with opposite changes in WQL20 and WQL30 and significantly differentially expressed were identified using the following criteria: [1] total reads $\geq 3,000$; [2] [treated/mock] ≥ 2 or [treated/mock] ≤ 0.5 in at least one stage.

gene *SPL* and negatively regulates cold tolerance in rice (Cui et al., 2015); sha-miR319 can silence *Gamyb-like1* genes that positively regulate cold tolerance in tomatoes (Shi et al., 2019). As a typical semi-arid tropical and subtropical plant, peanuts are highly sensitive to cold stress and may have evolved differential regulatory mechanisms to respond to cold stress. However, the function of small RNA-mediated gene silencing in peanut cold defense has not been reported. In this study, the transcription levels of miRNAs were analyzed between tolerant and sensitive lines with cold treatment in peanuts, and 407 known miRNAs and 143 novel peanut-specific miRNAs in total were identified. Our data showed that known miRNAs and novel miRNAs were involved in cold responses from eight libraries (**Supplementary Table S1**). Some of these miRNAs have been reported to be responsive to low temperatures in other plants. For example, specific members of the miR171 and miR396 families were also found to respond to cold treatment in peanuts, indicating that they are also conserved in dicotyledons. Likewise, some previously identified cold-induced conserved miRNAs, such as miR393 and miR402 in *Arabidopsis* (Sunkar and Zhu, 2004), were not found in our data, suggesting that they were unaffected by cold or that induction levels of certain miRNAs may be too low to be detected as significant in peanuts. Finally, through experimental verification, we found that six miRNAs might play key roles in peanut cold defense.

Based on miRNA expression profiling and degradome analysis combined with northern blotting and qRT-PCR validation, we found that several miRNA target pairs (miR160-*ARF*, miR482-*WDRL*, miR2118-*DR*, miR162-*DCL*, miR1511-*SRF*, miR1511-*SPIRAL1*, and miR396-*GRF*) might play key roles in the regulation of cold tolerance in peanuts. Specially, miR160-*ARF*, miR482-*WDRL*, and miR2118-*DR* showed opposite expression profiles to miR162-*DCL*, miR1511-*SRF*, miR1511-*SPIRAL1*, and miR396-*GRF* in tolerant and sensitive lines, indicating that they may be key factors leading to differences in cold tolerance.

miRNAs Mediated Positive Regulation Pathway in Peanut Cold Response

ahy-miR160, ahy-miR482, and ahy-miR2118 were upregulated in cold-tolerant line WQL30 but downregulated in cold-sensitive line WQL20 (**Figure 3**). According to previous reports, highly conserved miR160 with *ARF* targets responds to cold tolerance in maize, wheat, *Populus trichocarpa*, *Populus simonii* \times *P. nigra*, and wild banana (Lu et al., 2005; Liu et al., 2018; Zhou et al., 2019; Aydinoglu, 2020). Likewise, the target gene auxin response factor (*ARF*) not only regulates plant growth but also participates in abiotic stress responses, especially cold stress. For example, in *Arabidopsis*, *AtARF* genes are upregulated during cold acclimation (Hannah et al., 2005). Many *ARF* genes are repressed or induced in bananas (*Musa acuminata* L.) following cold, salt, and osmotic stress (Hu et al., 2015). In maize, 13 *ARFs* were upregulated in seedling leaves after cold stress (Aydinoglu, 2020). SLR/IAA14, a transcriptional repressor of auxin signaling, plays a crucial role in the integration of miRNAs in auxin and

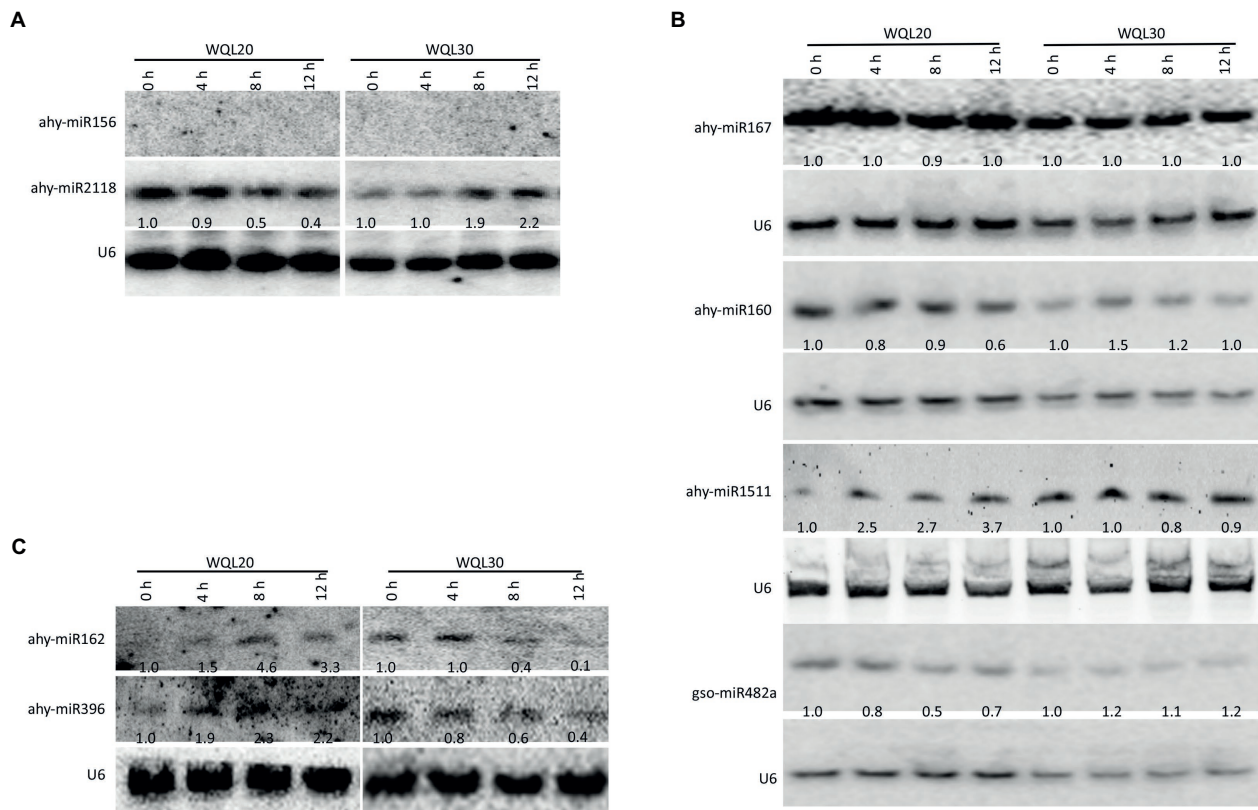


FIGURE 3 | Expression patterns of miRNAs upon low-temperature treatment. Four-leaf stage seedlings were treated with low temperature (6°C), and total RNA was extracted from leaves at the indicated time points. Detection of the peanut miRNAs using Northern blotting. A total of 100 µg of total RNA was loaded. RNA blots were hybridized with DNA oligonucleotide probes complementary to the indicated miRNAs. U6 was used as a loading control. Values below each section represent the relative abundance of miRNA normalized to U6.

cold responses in *Arabidopsis* roots (Aslam et al., 2020). In peanuts, the target genes of ahy-miR160 (*AhARF10* and *AhARF17*) also responded to cold stress (Figures 4, 5), and overexpression of ahy-miR160 enhanced cold tolerance in tobacco (Figure 6). These studies indicate that under cold stress conditions, upregulated ahy-miR160 may reduce ARF transcript levels, suppressing ARF-mediated auxin-responsive gene expression, leading to peanut growth and development attenuation, eventually enhancing cold stress tolerance in peanuts.

Studies have reported that miR2118 targets the protein family associated with disease resistance in response to biotic and abiotic stresses in many plants (de Vries et al., 2015; Wu et al., 2015; Cakir et al., 2016; Gao et al., 2016; Yang et al., 2020; Shi et al., 2021). In *Caragana intermedia*, cin-miR2118, a drought-resistant miRNA, positively affects drought stress tolerance. Overexpression of cin-miR2118 enhances drought tolerance in tobacco (Wu et al., 2015). miR2118 is a cold-responsive sRNA in wheat, banana, *Astragalus membranaceus*, and cassava (Xia et al., 2014; Song et al., 2017; Abia et al., 2019; Zhu et al., 2019). Similarly, our data confirmed that ahy-miR2118 was also a cold-responsive miRNA (Figure 3), and DR genes could also

be targeted by ahy-miR2118 via the anti-correlation between miR2118 and the corresponding target *AhDR* genes in peanuts (Figures 4, 5). Moreover, enhanced cold tolerance was also observed in tobacco plants overexpressing ahy-miR2118 (Figure 6), indicating that miR2118 might play an important role in plant resistance to abiotic stress. Under cold treatment, pdu-miR482 was differentially expressed between H (a cold-tolerant genotype) and Sh12 (a cold-sensitive genotype) in almonds (Karimi et al., 2016), respectively. Csn-miR482 was also significantly different between two tea plant cultivars, “Yingshuang” (YS, a cold-tolerant tea plant cultivar) and “Baiye 1” (BY, a cold-sensitive tea plant cultivar; Zhang et al., 2014). In peanuts, the repression in sensitive lines and weak induction in tolerant lines for ahy-miR482 was consistent with the findings of the ovary tissues of H genotype and Sh12 under 0°C treatment in almonds (Figure 3). ahy-miR482 targeted *AhWDRs* (Figure 4). *AhWDRs* belong to the WD40 protein family, and the large gene family of WD40 proteins is involved in a broad spectrum of crucial plant stress resistance processes. HOS15, a WD40-repeat protein, plays a role in gene activation/repression via histone modification during plant acclimation to low-temperature conditions (Sharma and Pandey, 2015). RNA-seq data analysis

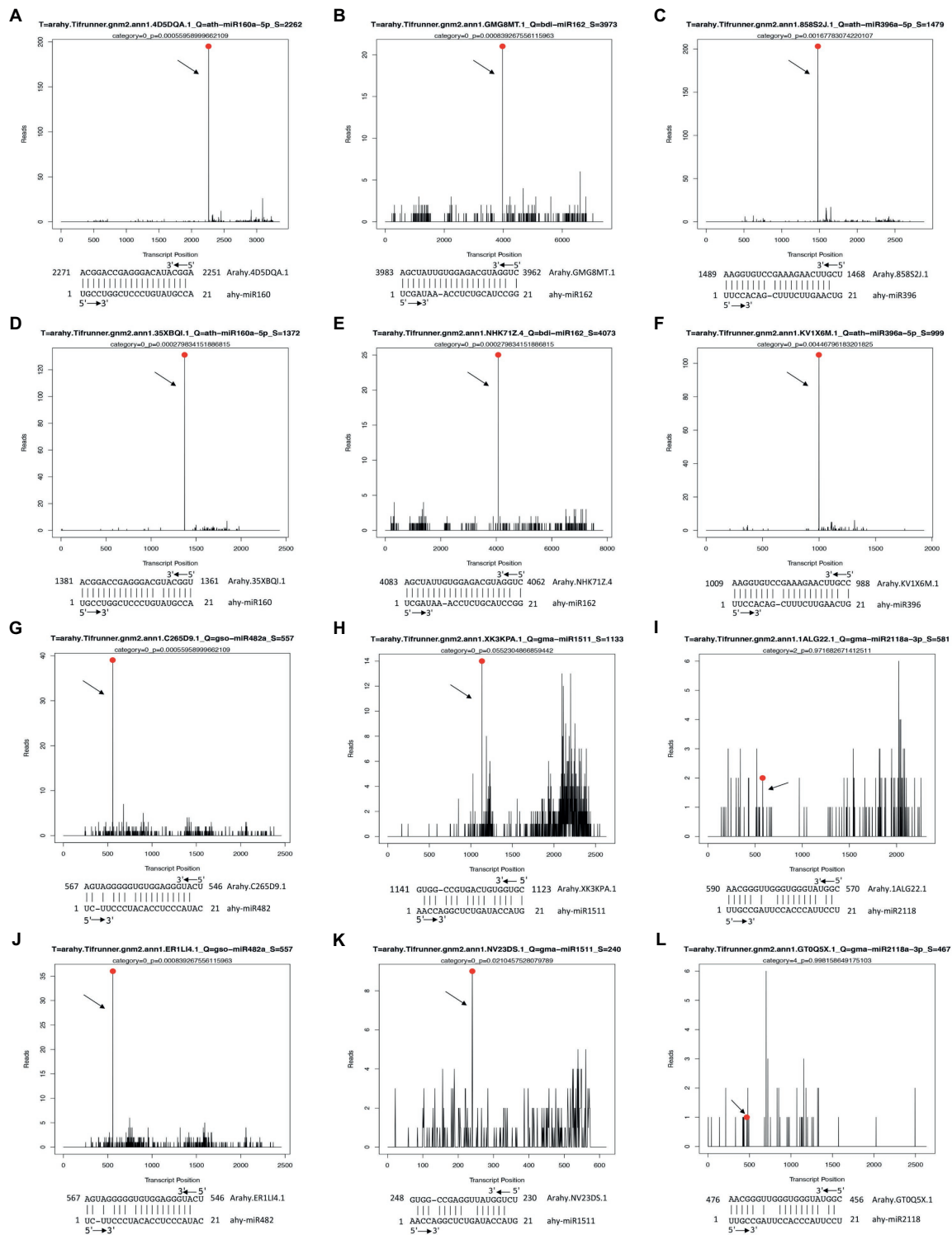


FIGURE 4 | Examples of T-plots of miRNA targets in two peanut RILs confirmed by degradome sequencing. The T-plots show the distribution of the degradome tags along the full length of the target mRNA sequence. The vertical red line indicates the cleavage site of each transcript and is also shown by an arrow. **(A,D)** The cleavage features in *AhARF10* (Arahy.4D5DQA.1) and *ARF17* (Arahy.35XBQI.1) mRNA by ahy-miR160 in DS1. **(B,E)** The cleavage features in *AhDCL6* (Arahy.GMG8MT.1) and *AhDCL16* (Arahy.NHK71Z.4) mRNA by ahy-miR162 in DS1. **(C,F)** The cleavage features in *AhGRF1* (Arahy.858S2J.1) and *AhGRF4* (Arahy.KV1X6M.1) mRNA by ahy-miR396 in DS1. **(G,J)** The cleavage features in *AhWDR1* (Arahy.C265D9.1) and *AhWDR2* (Arahy.ER1LI4.1) mRNA by ahy-miR482 in DS1. **(H,K)** The cleavage features in *AhSRF* (Arahy.XK3KPA.1) and *AhSP1L1* (Arahy.NV23DS.1) mRNA by ahy-miR1511 in DS1. **(I,L)** The cleavage features in *AhDR1* (Arahy.1ALG22.1) and *AhDR2* (Arahy.GTQ5X.1) mRNA by ahy-miR2118 in DS1.

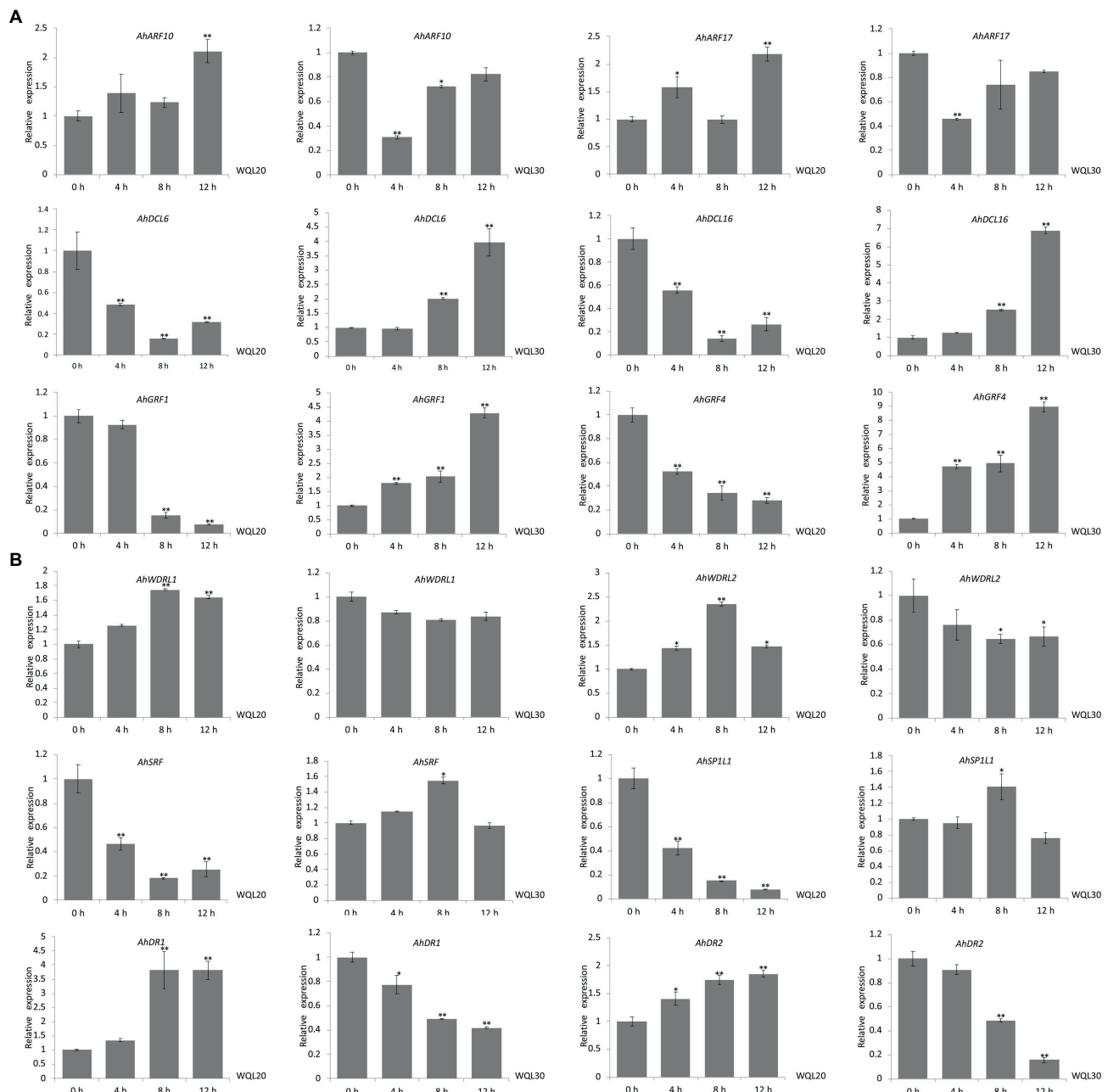


FIGURE 5 | Comparison of the relative expression levels of target genes in two RILs. **(A)** qRT-PCR analyses of target genes of ahy-miR160, ahy-miR162, and ahy-miR396 expression levels in two peanut RILs upon low-temperature (6°C) treatment at the indicated time points, respectively. **(B)** qRT-PCR analyses of target genes of ahy-miR482, ahy-miR1511, and ahy-miR2118 expression levels in two peanut RILs upon low-temperature (6°C) treatment at the indicated time points, respectively. Data are shown as means \pm SD ($n=3$). Student's t -test was used to determine the significance of differences between 0 HCT and the indicated time points. Asterisks indicate significant differences (* $p<0.05$ and ** $p<0.01$). Similar results were obtained from three biological replicates.

also showed that numerous *TaWD40s* are involved in responses to cold stress in wheat (Hu et al., 2018). Similarly, we found that *AhWDRL* genes are also involved in responses to cold stress in peanuts (Figure 5). Moreover, the phenomenon of enhanced cold tolerance was observed in tobacco by overexpressing ahy-miR482 (Figure 6). In brief, miR160-ARF,

miR482-WDRL, and miR2118-DR modules might mediate cold stress responses, and these three miRNAs may play roles in the positive regulation of peanut cold tolerance through related target genes in peanuts. Therefore, our data sheds light on the possibility of manipulating miRNAs to improve the tolerance of peanuts to cold stress.

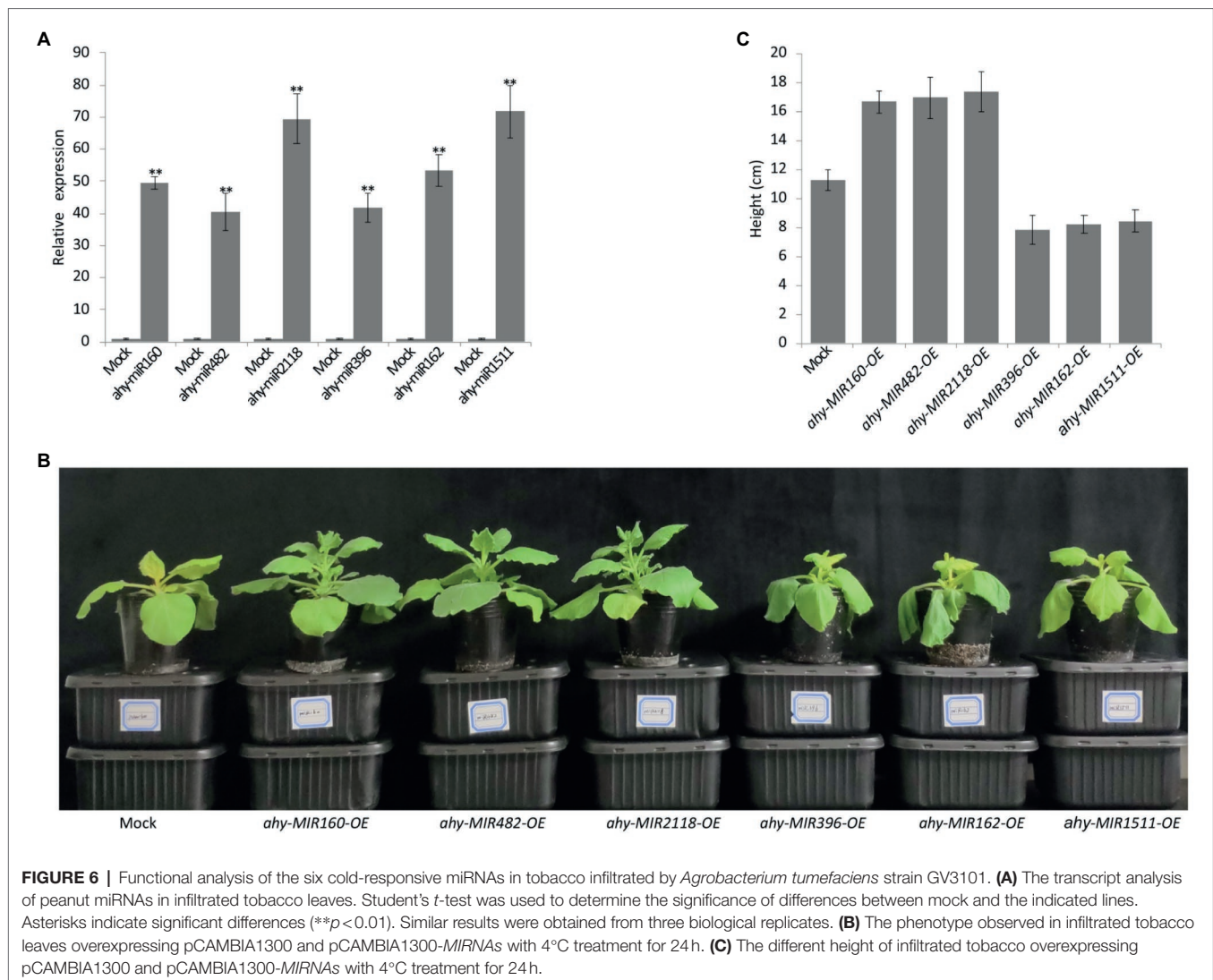


FIGURE 6 | Functional analysis of the six cold-responsive miRNAs in tobacco infiltrated by *Agrobacterium tumefaciens* strain GV3101. **(A)** The transcript analysis of peanut miRNAs in infiltrated tobacco leaves. Student's *t*-test was used to determine the significance of differences between mock and the indicated lines. Asterisks indicate significant differences (** $p < 0.01$). Similar results were obtained from three biological replicates. **(B)** The phenotype observed in infiltrated tobacco leaves overexpressing pCAMBIA1300 and pCAMBIA1300-MIRNAs with 4°C treatment for 24 h. **(C)** The different height of infiltrated tobacco overexpressing pCAMBIA1300 and pCAMBIA1300-MIRNAs with 4°C treatment for 24 h.

miRNAs Mediated Negative Regulation Pathway in Peanut Cold Response

In addition to the above-mentioned miRNAs, we found three other miRNAs with differential responses to cold stress in tolerant and sensitive lines: ahy-miR396, ahy-miR162, and ahy-miR1511, which were upregulated in the cold-sensitive line WQL20 but downregulated in the cold-tolerant line WQL30 (Figure 3). Overexpressing the three miRNAs weakened cold tolerance in tobacco, which means that they may play roles as negative regulators of cold tolerance in peanuts.

Previous studies have shown that miR396 mediates the silencing of target gene *GRFs* and participates in the regulation of the cold signal response in many plants, such as *Arabidopsis thaliana*, rice, soybean, tomato, poplar, wheat, and wild grapes (Lu et al., 2008; Tang et al., 2012; Zhang et al., 2014; Yan et al., 2016; Chen et al., 2019; Shi et al., 2019; Wang et al., 2019; Lantzouni et al., 2020). *GRFs* are a family of transcription factors that mediate the development of seeds, leaves, flowers, root growth, and other essential life processes in plants, such

as *Arabidopsis thaliana*, *Brassica napus*, *Glycine max* (soybean), *Solanum tuberosum* (potato), *Oryza sativa* (rice), and *Zea mays* (maize) (Kim et al., 2003; Omidbakhshfard et al., 2015). Additionally, *GRFs* are also the critical factors of cold signal transduction in plants, which can directly bind to the promoter region of the *CBF* gene to inhibit its transcription. For example, *Arabidopsis AtGRF7* can bind to the *gtcagg* cis element on the *AtCBF3* promoter to inhibit the expression of *AtCBF3* at the transcriptional level (Kim et al., 2012). Lantzouni et al. (2020) found that *AtCBF* genes (*AtCBF1*, *AtCBF2*, and *AtCBF3*) and the downstream *AtCOR* genes (*AtCOR15a*, *AtCOR15b*, and *AtCOR29d*) were significantly downregulated *in vivo* after overexpression of *GRF5*, and miR396/*GRF* module is a vital regulation mode for plants to balance growth, development, and cold response under cold stress. In peanuts, we also observed that ahy-miR396 can target *AhGRFs* and that overexpression of ahy-miR396 weakened cold tolerance in tobacco (Figures 4, 6), suggesting that it might help alleviate cold tolerance in peanuts through

the target *GRF* gene. However, there is no direct evidence for a relationship between *AhGRFs* and the CBF pathway in peanuts.

Studies have shown that miR1511 also is a cold-responsive sRNA in many plants. For example, pdu-miR1511 was differentially expressed between H and Sh12 in almond under cold treatment (Karimi et al., 2016). csi-miR1511 was also significantly different between YS and BY in tea plant (Zhang et al., 2014). mit-miR1511 showed remarkably higher levels at the cold stress temperatures (0°C) in the wild bananas (Liu et al., 2018). In agreement with previous reports, ahy-miR1511 showed significant differences between the tolerant and sensitive lines (Figure 3). It targets the *AhSRF* and *AhSPIRAL1* genes (Figures 4, 5). In *Arabidopsis*, *AtSPIRAL1* maintains microtubules (MT) stability and is closely associated with stress. For example, salt stress can lead to the UPS-dependent degradation of the MT-associated protein *AtSPIRAL1*. The degradation of *AtSPIRAL1* results in the depolymerization of MTs, followed by the formation of new MTs that are better adapted to osmotic stress (Wang et al., 2011). For another target gene *AhSRF*, recent evidence has shown that homologous gene *AtSRF6* is an identified regulator for *COR* in *Arabidopsis* (Wei et al., 2021). These reports suggest that the miR1511-*SRF* and miR1511-*SPIRAL1* modules may attenuate cold tolerance by mediating MT stability or by regulating the ICE-CBF-COR pathway. Another anti-correlated module in peanuts is ahy-miR162, which targets *AhDCL*. It has been reported that miR162 not only regulates the key factors of plant growth and development but also responds to low-temperature stress. In almonds, pdu-miR162 is upregulated in H and downregulated in Sh12 under cold stress (Karimi et al., 2016). This difference in expression in tolerant and sensitive varieties agreed with our data in peanuts (Figure 3), suggesting that miR162 is conserved and functional in plant cold tolerance. Pdu-miR162 directly targets *PdDCL1* for cleavage (Karimi et al., 2016). DCL proteins are key regulators of small RNA biogenesis (RNA interference, RNAi), and participate in the response to cold stress in plants. For example, rice, like peanuts, are low temperature-sensitive plants. Vyse et al. (2020) found that *OsDCL1* is induced by cold acclimation in rice. We also observed significant differences in the expression of *AhDCL6* and *AhDCL16* targeted by ahy-miR162 in tolerant and sensitive lines (Figures 4, 5), suggesting that the miR162-*DCL* module probably weakens cold stress tolerance in peanuts by feedback regulation of small RNA biogenesis.

CONCLUSION

In summary, cold-responsive miRNAs and candidate target genes were identified through integrated sRNA and degradome analysis during cold treatment of tolerant and sensitive lines. ahy-miR160, ahy-miR162, ahy-miR396, ahy-miR482, ahy-miR1511, and ahy-miR2118, as well as transcriptional factors, including *GRF*, *WDRL*, and *ARF*, or genes, such as *SRF*, *DCL*, and *SPIRAL*, were differentially expressed. They

might be the main contributors to the CBF and ARF pathways, protein kinase, small RNA biogenesis, and MT stability involved in cold tolerance of peanuts. These miRNAs may downregulate the expression of their target genes, which encode the regulatory and functional proteins involved in cold tolerance. These results increase our knowledge of sRNAs involved in the post-transcriptional regulation of cold tolerance and provide candidate genes for future functional analyses of cold tolerance-related signaling pathways in peanuts.

DATA AVAILABILITY STATEMENT

The raw data of the sRNA and degradome libraries are available from the NCBI Sequence Read Archive (SRA) under accession numbers SRR19241918, SRR19241919, SRR19241920, SRR19241921, SRR19241922, SRR19241923, SRR19241924, SRR19241925 and SRR19262409, SRR19262410, respectively.

AUTHOR CONTRIBUTIONS

DB and HJ designed the experiments. XZ, CR, YX, YT, HZ, NL, and CS performed the experiments. XZ, DB, and HJ performed the data analysis. XZ wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (31871662), the Open Project of Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, P. R. China (KF2021004), the China Agricultural Research System (CARS-13), the Cultivation Project of National Natural Science Foundation (YGJPY1901), the Doctoral Research Fund Project, Technology Innovation Research Project of the Shanxi Academy of Agricultural Sciences (YBSJJ2014), the Research Program Sponsored by State Key Laboratory of Sustainable Dryland Agriculture (in preparation), Shanxi Agricultural University (NO. 202105D121008), and the earmarked fund for Modern Agro-industry Technology Research System (2022-05).

ACKNOWLEDGMENTS

We are grateful to Huifang Jiang for kindly providing WQL20 and WQL30 seeds, and Editage (www.editage.cn) for English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.920195/full#supplementary-material>

Supplementary Figure S1 | Length distribution and abundance of the sequences. The lengths of the unique, valid reads ranged from 18 to 30 nucleotides (nt), and the 21–24 nt sequences were predominant in all libraries, with the 24 (nt) sequences being the most common.

Supplementary Figure S2 | Examples of T-plots of miRNA targets in two peanut RILs confirmed by degradome sequencing. The T-plots show the distribution of the degradome tags along the full length of the target mRNA sequence. The vertical red line indicates the cleavage site of each transcript and is also shown by an arrow. **(A,D)** The cleavage features in *AhARF10* (Arahy.4D5DQA.1) and *AhARF17* (Arahy.35XBQL.1) mRNA by ahy-miR160 in DS2. **(B,E)** The cleavage features in *AhDCL6* (Arahy.GMG8MT.1) and *AhDCL16* (Arahy.NHK71Z.4) mRNA by ahy-miR162 in DS2. **(C,F)** The cleavage features in *AhGRF1* (Arahy.858S2J.1) and *AhGRF4* (Arahy.KV1X6M.1) mRNA by ahy-miR396 in DS2.

Supplementary Figure S3 | Examples of T-plots of miRNA targets in two peanut RILs confirmed by degradome sequencing. The T-plots show the

distribution of the degradome tags along the full length of the target mRNA sequence. The vertical red line indicates the cleavage site of each transcript and is also shown by an arrow. **(A,D)** The cleavage features in *AhWDR1* (Arahy.C265D9.1) and *AhWDR2* (Arahy.ER1LJ4.1) mRNA by ahy-miR482 in DS2. **(B,E)** The cleavage features in *AhSRF* (Arahy.XK3KPA.1) and *AhSP1L1* (Arahy.NV23DS.1) mRNA by ahy-miR1511 in DS2. **(C,F)** The cleavage features in *AhDR1* (Arahy.1ALG22.1) and *AhDR2* (Arahy.GT0Q5X.1) mRNA by ahy-miR2118 in DS2.

Supplementary Table S1 | List of identified miRNAs differentially expressed in peanut with cold treatment.

Supplementary Table S2 | Statistics of degradome reads in DS1 and DS2.

Supplementary Table S3 | Differentially-expressed target genes between the DS1 and DS2 libraries.

Supplementary Table S4 | Primer information.

REFERENCES

- Abla, M., Sun, H. G., Li, Z. Y., Wei, C. X., Gao, F., Zhou, Y. J., et al. (2019). Identification of miRNAs and their response to cold stress in *Astragalus Membranaceus*. *Biomol. Ther.* 9:182. doi: 10.3390/biom9050182
- Addo-Quaye, C., Miller, W., and Axtell, M. J. (2009). CleaveLand: a pipeline for using degradome data to find cleaved small RNA targets. *Bioinformatics* 25, 130–131. doi: 10.1093/bioinformatics/btn604
- Aslam, M., Sugita, K., Qin, Y., and Rahman, A. (2020). Aux/IAA14 regulates microRNA-mediated cold stress response in Arabidopsis roots. *Int. J. Mol. Sci.* 21:8441. doi: 10.3390/ijms21228441
- Aydinoglu, F. (2020). Elucidating the regulatory roles of microRNAs in maize (*Zea mays* L.) leaf growth response to chilling stress. *Planta* 251:38. doi: 10.1007/s00425-019-03331-y
- Barrero-Gil, J., and Salinas, J. (2018). Gene regulatory networks mediating cold acclimation: the CBF pathway. *Adv. Exp. Med. Biol.* 1081, 3–22. doi: 10.1007/978-981-13-1244-1_1
- Cakir, O., Candar-Cakir, B., and Zhang, B. H. (2016). Small RNA and degradome sequencing reveals important microRNA function in *Astragalus chrysochlorus* response to selenium stimuli. *Plant Biotechnol. J.* 14, 543–556. doi: 10.1111/pbi.12397
- Chen, N., Yang, Q., Su, M., Pan, L. J., Chi, X. Y., Chen, M. N., et al. (2012). Cloning of six ERF family transcription factor genes from peanut and analysis of their expression during abiotic stress. *Plant Mol. Biol. Report.* 30, 1415–1425. doi: 10.1007/s11105-012-0456-0
- Chen, X., Jiang, L. R., Zheng, J. S., Chen, F. Y., Wang, T. S., Wang, M. L., et al. (2019). A missense mutation in Large Grain Size 1 increases grain size and enhances cold tolerance in rice. *J. Exp. Bot.* 70, 3851–3866. doi: 10.1093/jxb/erz192
- Cui, N., Sun, X., Sun, M., Jia, B., Duanmu, H., Lv, D., et al. (2015). Overexpression of osmiR156k leads to reduced tolerance to cold stress in rice (*Oryza Sativa*). *Mol. Breed.* 35:214. doi: 10.1007/s11032-015-0402-6
- de Vries, S., Kloesges, T., and Rose, L. E. (2015). Evolutionarily dynamic, but robust, targeting of resistance genes by the miR482/2118 gene family in the Solanaceae. *Genome Biol. Evol.* 7, 3307–3321. doi: 10.1093/gbe/evv225
- Gao, F., Wang, N., Li, H. Y., Liu, J. S., Fu, C. X., Xiao, Z. H., et al. (2016). Identification of drought-responsive microRNAs and their targets in *Ammopiptanthus mongolicus* by using high-throughput sequencing. *Sci. Rep.* 6:34601. doi: 10.1038/srep34601
- Gilmour, S. J., Zarka, D. G., Stockinger, E. J., Salazar, M. P., Houghton, J. M., and Thomashow, M. F. (1998). Low temperature regulation of the Arabidopsis CBF family of AP2 transcriptional activators as an early step in cold-induced COR gene expression. *Plant J.* 16, 433–442. doi: 10.1046/j.1365-313x.1998.00310.x
- Guo, J., Ren, Y. K., Tang, Z. H., Shi, W. P., and Zhou, M. X. (2019). Characterization and expression profiling of the ICE-CBF-COR genes in wheat. *PeerJ* 7:e8190. doi: 10.7717/peerj.8190
- Hannah, M. A., Heyer, A. G., and Hinch, D. K. (2005). A global survey of gene regulation during cold acclimation in *Arabidopsis thaliana*. *PLoS Genet.* 1:e26. doi: 10.1371/journal.pgen.0010026
- Hu, R., Xiao, J., Gu, T., Yu, X., Zhang, Y., Chang, J., et al. (2018). Genome-wide identification and analysis of WD40 proteins in wheat (*Triticum aestivum* L.). *BMC Genomics* 19:803. doi: 10.1186/s12864-018-5157-0
- Hu, W., Zuo, J., Hou, X., Yan, Y., Wei, Y., Liu, J., et al. (2015). The auxin response factor gene family in banana: genome-wide identification and expression analyses during development, ripening, and abiotic stress. *Front. Plant Sci.* 6:742. doi: 10.3389/fpls.2015.00742
- Jeon, J., and Kim, J. (2013). Cold stress signaling networks in Arabidopsis. *J. Plant. Biol.* 56, 69–76. doi: 10.1007/s12374-013-0903-y
- Jin, H. L. (2008). Endogenous small RNAs and antibacterial immunity in plants. *FEBS Lett.* 582, 2679–2684. doi: 10.1016/j.febslet.2008.06.053
- Karimi, M., Ghazanfari, F., Fadaei, A., Ahmadi, L., Shiran, B., Rabei, M., et al. (2016). The small-RNA profiles of almond (*Prunus dulcis* mill.) reproductive tissues in response to cold stress. *PLoS One* 11:e0156519. doi: 10.1371/journal.pone.0156519
- Katiyar-Agarwal, S., and Jin, H. (2007). Discovery of pathogen-regulated small RNAs in plants. *Methods Enzymol.* 427, 215–227. doi: 10.1016/S0076-6879(07)27012-0
- Kim, J. H., Choi, D., and Kende, H. (2003). The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. *Plant J.* 36, 94–104. doi: 10.1046/j.1365-313X.2003.01862.x
- Kim, J. S., Mizoi, J., Kidokoro, S., Maruyama, K., Nakajima, J., Nakashima, K., et al. (2012). Arabidopsis growth regulating factor7 functions as a transcriptional repressor of abscisic acid- and osmotic stress-responsive genes, including DREB2A. *Plant Cell* 24, 3393–3405. doi: 10.1105/tpc.112.100933
- Lantzouni, O., Alkofer, A., Falter-Braun, P., and Schwegheimer, C. (2020). GROWTH REGULATING FACTORS interact with DELLAs and regulate growth in cold stress. *Plant Cell* 32, 1018–1034. doi: 10.1105/tpc.19.00784
- Liu, W. H., Cheng, C. Z., Chen, F. L., Ni, S. S., Lin, Y. L., and Lai, Z. X. (2018). High-throughput sequencing of small RNAs revealed the diversified cold-responsive pathways during cold stress in the wild banana (*Musa itinerans*). *BMC Plant Biol.* 18:308. doi: 10.1186/s12870-018-1483-2
- Lu, S., Sun, Y. H., and Chiang, V. L. (2008). Stress-responsive microRNAs in Populus. *Plant J.* 55, 131–151. doi: 10.1111/j.1365-313X.2008.03497.x
- Lu, S., Sun, Y. H., Rui, S., Clark, C., and Chiang, V. L. (2005). Novel and mechanical stress-responsive microRNAs in *Populus trichocarpa* that are absent from Arabidopsis. *Plant Cell* 17, 2186–2203. doi: 10.1105/tpc.105.033456
- Niu, D. D., Zhang, X., Song, X. O., Wang, Z. H., Li, Y. Q., Qiao, L. L., et al. (2018). Deep sequencing uncovers rice long siRNAs and its involvement in immunity against *Rhizoctonia solani*. *Phytopathology* 108, 60–69. doi: 10.1094/Phyto-03-17-0119-R
- Omidbakhshfar, M. A., Proost, S., Fujikura, U., and Mueller-Roeber, B. (2015). Growth-Regulating Factors (GRFs): A small transcription factor family with important functions in plant biology. *Mol. Plant* 8, 998–1010. doi: 10.1016/j.molp.2015.01.013

- Park, S., Lee, C. M., Doherty, C. J., Gilmour, S. J., Kim, Y. S., and Thomashow, M. F. (2015). Regulation of the Arabidopsis CBF regulon by a complex low temperature regulatory network. *Plant J.* 82, 193–207. doi: 10.1111/tpj.12796
- Pruthi, V., Narasimhan, R., and Nataraja, K. N. (2014). Simultaneous expression of abiotic stress responsive transcription factors, AtDREB2A, AtHB7 and AtABF3 improves salinity and drought tolerance in peanut (*Arachis hypogaea* L.). *PLoS One* 9:e111152. doi: 10.1371/journal.pone.0111152
- Sharma, M., and Pandey, G. K. (2015). Expansion and function of repeat domain proteins during stress and development in plants. *Front. Plant Sci.* 6:1218. doi: 10.3389/fpls.2015.01218
- Shi, Y., Ding, Y., and Yang, S. (2018). Molecular regulation of CBF signaling in cold acclimation. *Trends Plant Sci.* 23, 623–637. doi: 10.1016/j.tplants.2018.04.002
- Shi, X., Jiang, F., Wen, J., and Wu, Z. (2019). Overexpression of *Solanum habrochaites* microRNA319d (sha-miR319d) confers chilling and heat stress tolerance in tomato (*S. lycopersicum*). *BMC Plant Biol.* 19:214. doi: 10.1186/s12870-019-1823-x
- Shi, R., Jiao, W., Yun, L., Zhang, Z., Zhang, X., Wang, Q., et al. (2021). Utilization of transcriptome, small rna, and degradome sequencing to provide insights into drought stress and rewatering treatment in *Medicago ruthenica*. *Front. Plant Sci.* 12:675903. doi: 10.3389/fpls.2021.675903
- Song, G. Q., Zhang, R. Z., Zhang, S. J., Li, Y. L., Gao, J., Han, X. D., et al. (2017). Response of microRNAs to cold treatment in the young spikes of common wheat. *BMC Genomics* 18:212. doi: 10.1186/s12864-017-3556-2
- Sunkar, R., and Zhu, J. K. (2004). Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. *Plant Cell* 16, 2001–2019. doi: 10.1105/tpc.104.022830
- Tang, Z., Zhang, L., Xu, C., Yuan, S., Zhang, F., Zheng, Y., et al. (2012). Uncovering small RNA-mediated responses to cold stress in a wheat thermosensitive genic male-sterile line by deep sequencing. *Plant Physiol* 159, 721–738. doi: 10.1104/pp.112.196048
- Upadhyaya, H. D., Rodomiro, O., Paula, J. B., and Sube, S. (2003). Development of a groundnut core collection using taxonomical, geographical and morphological descriptors. *Genet. Resour. Crop. Ev.* 50, 139–148. doi: 10.1007/s10722-003-6111-8
- Vyse, K., Faivre, L., Romich, M., Pagter, M., Schubert, D., Hinch, D. K., et al. (2020). Transcriptional and post-transcriptional regulation and transcriptional memory of chromatin regulators in response to low temperature. *Front. Plant Sci.* 11:39. doi: 10.3389/fpls.2020.00039
- Wang, D. Z., Jin, Y. N., Ding, X. H., Wang, W. J., Zhai, S. S., Bai, L. P., et al. (2017). Gene regulation and signal transduction in the ICE-CBF-COR signaling pathway during cold stress in plants. *Biochemistry* 82, 1103–1117. doi: 10.1134/S0006297917100030
- Wang, P., Yang, Y., Shi, H., Wang, Y., and Ren, F. (2019). Small RNA and degradome deep sequencing reveal respective roles of cold-related microRNAs across Chinese wild grapevine and cultivated grapevine. *BMC Genomics* 20:740. doi: 10.1186/s12864-019-6111-5
- Wang, S., Kurepa, J., Hashimoto, T., and Smalle, J. A. (2011). Salt stress-induced disassembly of Arabidopsis cortical microtubule arrays involves 26S proteasome-dependent degradation of SPIRAL1. *Plant Cell* 23, 3412–3427. doi: 10.1105/tpc.111.089920
- Wei, X. S., Liu, S., Sun, C., Xie, G. S., and Wang, L. Q. (2021). Convergence and divergence: signal perception and transduction mechanisms of cold stress in Arabidopsis and rice. *Plants-Basel* 10:1864. doi: 10.3390/plants10091864
- Wu, B. F., Li, W. F., Xu, H. Y., Qi, L. W., and Han, S. Y. (2015). Role of cin-miR2118 in drought stress responses in *Caragana intermedia* and tobacco. *Gene* 574, 34–40. doi: 10.1016/j.gene.2015.07.072
- Xia, J., Zeng, C. Y., Chen, Z., Zhang, K. V., Chen, X., Zhou, Y. F., et al. (2014). Endogenous small-noncoding RNAs and their roles in chilling response and stress acclimation in cassava. *BMC Genomics* 15:634. doi: 10.1186/1471-2164-15-634
- Xiao, F. J., and Song, L. C. (2011). Analysis of extreme low-temperature events during the warm season in Northeast China. *Na Hazards* 58, 1333–1344. doi: 10.1007/s11069-011-9735-6
- Yan, J., Zhao, C., Zhou, J., Yang, Y., Wang, P., Zhu, X., et al. (2016). The miR165/166 mediated regulatory module plays critical roles in ABA homeostasis and response in Arabidopsis thaliana. *PLoS Genet.* 12:e1006416. doi: 10.1371/journal.pgen.1006416
- Yang, C., Li, D., Mao, D., Liu, X., Ji, C., Li, X., et al. (2013). Overexpression of microRNA319 impacts leaf morphogenesis and leads to enhanced cold tolerance in rice (*Oryza sativa* L.). *Plant Cell Environ.* 36, 2207–2218. doi: 10.1111/pce.12130
- Yang, Z. M., Zhu, P. P., Kang, H., Liu, L., Cao, Q. H., Sun, J., et al. (2020). High-throughput deep sequencing reveals the important role that microRNAs play in the salt response in sweet potato (*Ipomoea batatas* L.). *BMC Genomics* 21:164. doi: 10.1186/s12864-020-6567-3
- Zhang, X., Bao, Y. L., Shan, D. Q., Wang, Z. H., Song, X. N., Wang, Z. Y., et al. (2018). *Magnaporthe oryzae* induces the expression of a microRNA to suppress the immune response in rice. *Plant Physiol.* 177, 352–368. doi: 10.1104/pp.17.01665
- Zhang, H., Dong, J., Zhao, X., Zhang, Y., Ren, J., Xing, L., et al. (2019). Research progress in membrane lipid metabolism and molecular mechanism in peanut cold tolerance. *Front. Plant Sci.* 10:838. doi: 10.3389/fpls.2019.00838
- Zhang, M., Liu, W., Bi, Y. P., and Wang, Z. Z. (2009). Isolation and identification of PNDREB1: a new DREB transcription factor from peanut (*Arachis hypogaea* L.). *Acta Agron. Sin.* 35, 1973–1980. doi: 10.3724/SPJ.1006.2009.01973
- Zhang, Y., Zhu, X. J., Chen, X., Song, C. N. A., Zou, Z. W., Wang, Y. H., et al. (2014). Identification and characterization of cold-responsive microRNAs in tea plant (*Camellia sinensis*) and their targets using high-throughput sequencing and degradome analysis. *BMC Plant Biol.* 14:271. doi: 10.1186/s12870-014-0271-x
- Zhou, B., Kang, Y., Leng, J., and Xu, Q. (2019). Genome-wide analysis of the miRNA-mRNAs network involved in cold tolerance in *Populus simonii* x *P. nigra*. *Genes-Basel* 10:430. doi: 10.3390/genes10060430
- Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029
- Zhu, H., Zhang, Y., Tang, R., Qu, H., Duan, X., and Jiang, Y. (2019). Banana sRNAome and degradome identify microRNAs functioning in differential responses to temperature stress. *BMC Genomics* 20:33. doi: 10.1186/s12864-018-5395-1
- Zhou, X., Wang, G., Sutoh, K., Zhu, J. K., and Zhang, W. (2008). Identification of cold-inducible microRNAs in plants by transcriptome analysis. *BBA-Gene Regul. Mech.* 1779, 780–788. doi: 10.1016/j.bbagr.2008.04.005

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Ren, Xue, Tian, Zhang, Li, Sheng, Jiang and Bai. 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(s) 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 Signatures in Chinese Sorghum Reveals Its Unique Liquor-Making Properties

Li Yi Zhang^{1†}, Yanqing Ding^{1†}, Jianxia Xu¹, Xu Gao¹, Ning Cao¹, Kuiying Li², Zhou Feng^{1,2}, Bing Cheng¹, Lengbo Zhou¹, Mingjian Ren², Xiaochun Lu³, Zhigui Bao⁴, Yuezhi Tao⁵, Zhanguo Xin^{6*} and Guihua Zou^{5*}

¹ Guizhou Institute of Upland Crops, Guizhou Academy of Agricultural Sciences, Guiyang, China, ² College of Agriculture, Guizhou University, Guiyang, China, ³ Institute of Sorghum Research, Liaoning Academy of Agricultural Sciences, Shenyang, China, ⁴ Shanghai OE Biotech Co., Ltd., Shanghai, China, ⁵ Institute of Crop and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, ⁶ Plant Stress and Germplasm Development Unit, Cropping Systems Research Laboratory, USDA-ARS, Lubbock, TX, United States

OPEN ACCESS

Edited by:

Chuanzhi Zhao,
Shandong Academy of Agricultural
Sciences, China

Reviewed by:

Weihua Qiao,
Institute of Crop Sciences (CAAS),
China
Feng Yu,
Hubei University, China

*Correspondence:

Li Yi Zhang
lyzhang1997@hotmail.com
Zhanguo Xin
zhanguo.xin@usda.gov
Guihua Zou
zouguihuazw@163.com

[†] These authors have contributed
equally to this work

Specialty section:

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

Received: 19 April 2022

Accepted: 20 May 2022

Published: 09 June 2022

Citation:

Zhang L, Ding Y, Xu J, Gao X,
Cao N, Li K, Feng Z, Cheng B,
Zhou L, Ren M, Lu X, Bao Z, Tao Y,
Xin Z and Zou G (2022) Selection
Signatures in Chinese Sorghum
Reveals Its Unique Liquor-Making
Properties.
Front. Plant Sci. 13:923734.
doi: 10.3389/fpls.2022.923734

Chinese sorghum (*S. bicolor*) has been a historically critical ingredient for brewing famous distilled liquors ever since Yuan Dynasty (749 ~ 652 years BP). Incomplete understanding of the population genetics and domestication history limits its broad applications, especially that the lack of genetics knowledge underlying liquor-brewing properties makes it difficult to establish scientific standards for sorghum breeding. To unravel the domestic history of Chinese sorghum, we re-sequenced 244 Chinese sorghum lines selected from 16 provinces. We found that Chinese sorghums formed three distinct genetic sub-structures, referred as the Northern, the Southern, and the Chishui groups, following an obviously geographic pattern. These sorghum accessions were further characterized in liquor brewing traits and identified selection footprints associated with liquor brewing efficiency. An importantly selective sweep region identified includes several homologous genes involving in grain size, pericarp thickness, and architecture of inflorescence. Our result also demonstrated that pericarp strength rather than grain size determines the ability of the grains to resist repeated cooking during brewing process. New insight into the traits beneficial to the liquor-brewing process provides both a better understanding on Chinese sorghum domestication and a guidance on breeding sorghum as a multiple use crop in China.

Keywords: sorghum, selective sweep, population structure, genome-wide association study (GWAS), liquor-brewing related traits

INTRODUCTION

Sorghum [*S. bicolor* (L.) Moench, $2n = 2x = 20$] is the fifth largest and widely adaptable crop in the world. It is cultivated in more than 100 countries and regions around the world. It is mainly used for food, feed, fiber, brewing, and bioenergy. Modern sorghum, like other domesticated crops, has been improved as a consequence of long-term artificial selection (Mace et al., 2013; Smith et al., 2019). The diverse distribution and extensive morphological variation observed in multiple sorghum species imply a remarkable history of divergence and evolution. The northeast quadrant of Africa is believed to be the primary origin and diversity center of sorghum. Shortly after its early domestication around 6,000 years BP (Before present), sorghum spread to India

(Doggett, 1988; Dahlberg and Wasylikowa, 1996; Winchell et al., 2017). The Indian subcontinent is considered as the secondary center of origin of sorghum, and the earliest domesticated sorghum appeared in the late Harappan, approximately 4,000–3,700 years BP (Fuller, 2003; Boivin and Fuller, 2009). Chinese sorghum, the Kaoliangs, has a cultivation history of ~2,200 years with morphologies characteristic of bicolor type. It was inferred as the early bicolor variants introduced from northeast India (Mann et al., 1983; Doggett, 1988). It is widely believed that the domesticated sorghum was transmitted eastward from India to China either *via* land or sea routes (Fuller, 2003; Boivin and Fuller, 2009). Recently, molecular genetics studies suggested that Chinese sorghum may be of African origin (Li et al., 2010; Zhang et al., 2011). This question remains unsolved based on the current evidence from morphology, agronomy, and even genomic features.

In China, sorghum has been a key ingredient for brewing liquor for ~750 years (Shinoda, 1958; Han, 2012). Ancient Chinese, as known in the world, is the first to master the knowledge of using sorghum to brew distilled liquor (Awika and Rooney, 2004; Zhao, 2019b). Currently, sorghum is widely cultivated across agroclimatic zones from the south to the north in China, and more than 80% of sorghum grain is used to brew liquor (Chen et al., 2019). Techniques to brew distilled liquor were developed in the Tang and Song Dynasties (1,402–749 years BP), and commercialized in the Yuan Dynasty (749–652 years BP) (Joseph, 2000). Almost all Chinese famous brands of distilled liquors, such as Moutaijiu and Langjiu (Moutai-aroma liquor), Luzhoulaojiao, Wuliangye (strong-aroma liquor), and Fenjiu (light-aroma liquor), etc., are brewed with sorghum grain as a key ingredient. The first two flavored liquors are mainly from southwest region, a principal area producing liquors since ancient times. The liquor-making processes of different flavor liquors have different requirements for the physical properties of sorghum grains. The Maotai-aroma liquor produced in the Chishui River Basin in the Southwest has a complex and special brewing process, which requires nine times of steaming and boiling, eight times of fermentation, and seven times of liquor extraction, hence, the sorghum grains must be able to resist to the repeatedly steaming and stirring. The brewing process of other flavored liquors only requires one or two times of cooking to extract the liquor, which desires sorghum grains that can be cooked and broken easily (Sun, 2019). This remarkably domesticated history by human-intervention raises the question to what extent the artificial and adaptive selections for liquor-brewing shaped the local Chinese sorghum population.

High-throughput resequencing technology, combined with comprehensive statistical analysis methods, provides a powerful tool to uncover the mysteries of plant domestication. Based on deep-coverage of whole-genome sequence of accessions from various regions, the origin and domestication history of several crops have recently been uncovered, including ancient sorghum (*S. bicolor ssp. bicolor* (L.) Moench.) (Smith et al., 2019), Tibetan barley (*Hordeum vulgare* L., qingke) (Zeng et al., 2018), castor (*Ricinus communis*) (Fan et al., 2019; Xu et al., 2021), pear (*Pyrus*) (Wu et al., 2018), chickpea (*Cicer arietinum* L.) (Varshney et al., 2019), etc. Therefore, we investigated the data from

whole genome resequencing of 333 sorghum accessions (244 sequenced in this study plus 89 genomes available publicly) to further explore the domestication history and genome-wide selective footprints on Chinese sorghum population. In addition, we described a genome wide map of SNP variation and traced patterns of sorghum diffusion to diverse agroclimatic regions. Our study demonstrated that Chinese sorghum was first introduced to south China from Indian and gradually spread to the north China. Furthermore, we identified genes underlying natural variation in traits related to baijiu-making process.

MATERIALS AND METHODS

Plant Material and Phenotyping

A worldwide sorghum collection of 333 accessions was used in this study, of which a panel of 244 was mainly composed of Chinese sorghum from the Center for Crop Germplasm Resources, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (Beijing, China), and Institute of Upland Crops, Guizhou Academy of Agricultural Science (Guiyang, China). Another panel of 89 accessions included 38 wild sorghum and 51 landraces/cultivars. Wild sorghums were mostly from Africa, while cultivated sorghums were from Africa and Asia, maintained in the Institute of Botany, Chinese Academy of Sciences (Beijing, China).

From 2018 to 2020, the panel of 244 accessions were grown in Guiyang, Guizhou Province; Lingshui and Ledong, Hainan Province; and Hangzhou, Zhejiang Province in China. Five traits related with brewing properties were thousand grain weight (TGW), grain length (GL), grain width (GW), pericarp thickness (PT) and testa thickness (TT). TGW, GL, and GW were measured by a Crop Grain Appearance Quality Scanning Machine (SC-E, Wanshen Technology Company, Hangzhou, China). PT and TT were measured by a Digital Microscope (VHX-7000, KEYENCE Technology Company, Shanghai, China).

DNA Extraction and Sequencing

Genomic DNAs of 244 accessions were extracted following the cetyltrimethylammonium bromide (CTAB) method and quantified by Qubit® 2.0 fluorescent meter (Invitrogen, Carlsbad, United States). The quality of the DNA was determined by electrophoresis in 0.8% agarose gel running at 100 V for 40 min. The DNA fragments around 350 bp were randomly generated by Covaris ultrasonic crushing apparatus. The constructed library was used for paired-end (PE150) sequencing on Illumina HiSeq 4000 sequencing platform by Beijing Novogene technology co., Ltd. (Beijing, China). The 89 public sequencing data for African and Asian sorghums were obtained from SorGSD (Zhang et al., 2018)¹.

Reads Mapping and Variants Calling

The raw paired-end reads were trimmed and filtered with a sliding window of size 4, with average Phred score scale of 20 within the window using fastp (version: 0.20.0)

¹<http://bigd.big.ac.cn/sorgsd>

(Chen et al., 2018). The clean reads of 333 accessions were mapped to the *Sorghum bicolor* genome (Phytozome v12²) (Goodstein et al., 2012) using bwa-mem (version: 0.7-17) (Li and Durbin, 2009) with default parameters. After alignment, Picard tools (version: 2.18.17³) were used to remove PCR duplicates according to the mapping coordinates.

The variation detection followed the best practice workflow recommended by Genome Analysis Toolkit (GATK4, version 3.8.1) (McKenna et al., 2010). In brief, the variants were called for each accession by the GATK HaplotypeCaller. A joint genotyping step for comprehensive variations union was performed on the gVCF files. Two subsets of the sorghum SNPs were defined using the following filtering criteria: (1) the basic set of 3,298,433 SNPs were created with exclusion of non bi-allelic, > 20% missing calls and MAF < 5%; (2) a core SNP set of 526,946 SNPs derived from the basic SNP set using a two-step linkage disequilibrium pruning procedure with PLINK (version: 1.9) (Purcell et al., 2007) with parameters “-indep-pairwise 50 5 0.5.”

SNPs and Indels were annotated according to the BTx623 genome using the SNPeff6 (version: 4.3T) (Cingolani et al., 2012). SNP and InDel densities across each chromosome were counted with 500 kb sliding window using VCFtools (version: 0.1.17) software (Danecek et al., 2011).

Phylogenetic and Population Analyses

A phylogenetic tree was constructed from the core SNP set by using the neighbor-joining method in the program PHYLIP (version: 3.697⁴), and IBS distance matrices were calculated using PLINK. The resulting phylogenetic tree was visualized using the online tool iTOL (version: 4.0) (Letunic and Bork, 2019).

Principal Components Analysis (PCA) was performed with the smartPCA program from EIGENSOFT package (version: 6.1.4) (Price et al., 2006) with the core SNP set, and the first three eigenvectors were plotted by scatterplot3d R package. Population structure was inferred using the fastSTRUCTURE (version: 1.0) program (Raj et al., 2014). To explore the convergence of individuals, we predefined the number of genetic clusters K from 2 to 12 and ran the cross-validation error (CV) procedure. fastSTRUCTURE was then run again on the whole core SNP set 10 times with varying random seeds; the Q-matrices were aligned using pong13 software and clustered based on similarity. Then, the matrices belonging to the largest cluster were averaged to produce the final matrix of admixture proportions.

Global and chromosomal r^2 for each sorghum group were evaluated using PopLDdecay (version: v3.41⁵) (Zhang et al., 2019) with default setting.

Genome Scanning for Selective Sweep Signals

We performed a genetic differentiation (*Fst*), nucleotide diversity (π) and Tajima'D based cross approach to investigate the

selection signals across the whole genome. A 10 kb sliding window with 10 kb step approach was applied to quantify *Fst*, π and Tajima'D by using VCFtools software (Danecek et al., 2011). The candidates that meet both top 5% of the three tests were selected as selective signals.

Genome-Wide Association Study

The association analysis was performed with the GLM and Blink statistical methods implemented in Genome Association and Prediction Integrated Tool package (GAPIT version: 3.0)⁶. The first three PCs derived from whole-genome SNPs were used as fixed effects in the mixed model to correct for stratification. The random effect was estimated from the groups clustered based on the kinship among all accessions. We defined the whole-genome significant cutoff with the adjusted Bonferroni test threshold, which was set as $P = -\log_{10}(0.05/2,015,850) = 7.6$. LD analysis and inference of the haplo-type blocks containing peak SNPs to document the potential candidate genes responsible for traits were performed in LDBlockShow (version: 1.39) (Dong et al., 2021).

Sequence Analysis of the Candidate Gene

The candidate gene (Sobic.002G228600) from 20 accessions with extreme phenotypes were amplified and sequenced by Sanger sequencing technology in Platinum Biotechnology Co., Ltd. (Shanghai, China). The raw sequences were assembled using DNASTAR Lasergene (version: 7.10) and alignments for the sequence dataset were performed by MAFFT (version: 7.48) to identify the variations. The ExpASY translate tool⁷ was used to convert nucleotide sequence into amino acid sequence, the sequencing results are shown in **Supplementary Figure 4**. The primers1 information: 5'-TTGGACAGTGTGGACTCTC-3' and 5'-CAGGAAGTAGGGATGGATCA-3', the primers2 information: 5'-TCGTCGTGTCGGTGAATAC-3' and 5'-TGTAACCTCAGCACGCAACAA-3'.

RNA-Seq and Data Analysis

Two sorghum cultivars, “654” and “LTR108” were chosen for RNA sequencing (RNA-seq). Young panicles samples were collected at the S1, S2, and S3 stages, which corresponded to less-than 5 cm in length, 10–15 cm in length, and 5 days after heading, respectively. Total RNA was isolated using a protocol reported previously (Zou et al., 2020). A total of 1 μ g RNA was used for library construction with an NEBNext®Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) according to the manufacturer's instructions. RNA sequencing was performed on NovaSeq 6000 platform. Clean reads were mapped to the sorghum reference genomes V3.1.1 on Phytozome database⁸ using Hisat2 (version: 2.2.1). The number of reads mapped to each gene was counted using HTSeq software (version: 0.9.1). Gene expression based on fragments per kilo-base of exon per million fragments mapped (FPKM) was next obtained

²<https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Phytozome>

³<http://broadinstitute.github.io/picard/>

⁴<http://evolution.genetics.washington.edu/phytip.html>

⁵<https://github.com/BGI-shenzhen/PopLDdecay>

⁶<https://www.maizegenetics.net/gapit>

⁷<http://web.expasy.org/translate/>

⁸<https://phytozome.jgi.doe.gov>

through Cufflinks (version: 2.2.1). Each sample consisted of three biological replicates.

Experiment of Resistance to Cooking

We performed the cooking experiment on 16 glutinous sorghums whose grain size and shape, thickness of pericarp and testa were distinctly dissimilar. A volume of the grains (V_0) with weight of 10 g of sorghum grains were measured by the drainage method in a 150 mL beaker. Beakers of glutinous and non-glutinous sorghum varieties were filled with 85°C water on standing in a thermostat for 8 and 10 h, respectively. Subsequently, the grain volume (V_t) was measured once again without the supernatant, and the expansion rate was thus obtained [Expansion rate = $(V_t - V_0)/V_0 \times 100\%$]. The crack rate was observed and counted from 100 random grains (Liu et al., 2012).

RESULTS

Whole Genome Resequencing of Chinese Sorghum

To investigate the population structure of Chinese sorghum accessions, we sequenced 244 sorghum landraces/cultivars, which were collected extensively from 16 sorghum planting provinces across China (151 from North China, 91 from South China, and 2 from abroad). To better address the domestication history of sorghum in China and identify the different genomic regions under domestication, we also supplemented our sequencing data with 89 publicly available sorghum genome sequences, including one accession of *Sorghum propinquum*, a wild relative of sorghum (the same subgenus with *S. bicolor*), 37 wild sorghum mainly from Africa, and 51 cultivated sorghum from Africa and Asia (Zhang et al., 2018). In total, a collection of 333 sorghum accessions was analyzed in the present study (Figure 1A and Supplementary Table 1).

The whole-genome resequencing generated a total of ~12.04 G paired-end reads 150 bp in length, with an average sequencing depth of $\sim 7.04 \times$ and an average genome coverage of ~86.8% (Supplementary Table 2). After mapping to the sorghum reference genome BTx623 (Paterson et al., 2009; McCormick et al., 2018) and single nucleotide polymorphism (SNP) calling, we obtained 3,298,433 high-quality SNPs and 429,794 InDels (insertions-deletions) from the 333 sorghum accessions. Among these variations, 203,278 (4.2%) SNPs were in coding regions, with a non-synonymous-to-synonymous substitution ratio of 1 (86,150 and 87,333 for non-synonymous and synonymous, respectively).

Genetic Structuration, Phylogenetic and Principal Component Analysis

The population structure of the whole data set ($N = 333$) was analyzed by *fastSTRUCTURE* with K ranging from 2 to 12 (Figure 1B). The barplots were inspected visually to identify the well delimited clusters that could be biologically relevant. Although the optimal number of clusters is inferred by the K selector as 5, the finest biologically relevant population structure

was identified at $K = 10$. The number of main clusters remained as six, including the wild relatives, cultivated sorghum in Africa, Southern Asia, Southern China, Northern China, and Chishui in Southwest China, which were apparently consistent with the evolutionary history of sorghum and corresponded to the species and/or geographical regions (Fuller, 2003; Boivin and Fuller, 2009). We therefore considered these six clusters as the most relevant genetic structure and estimated the sorghum individual membership at $K = 10$.

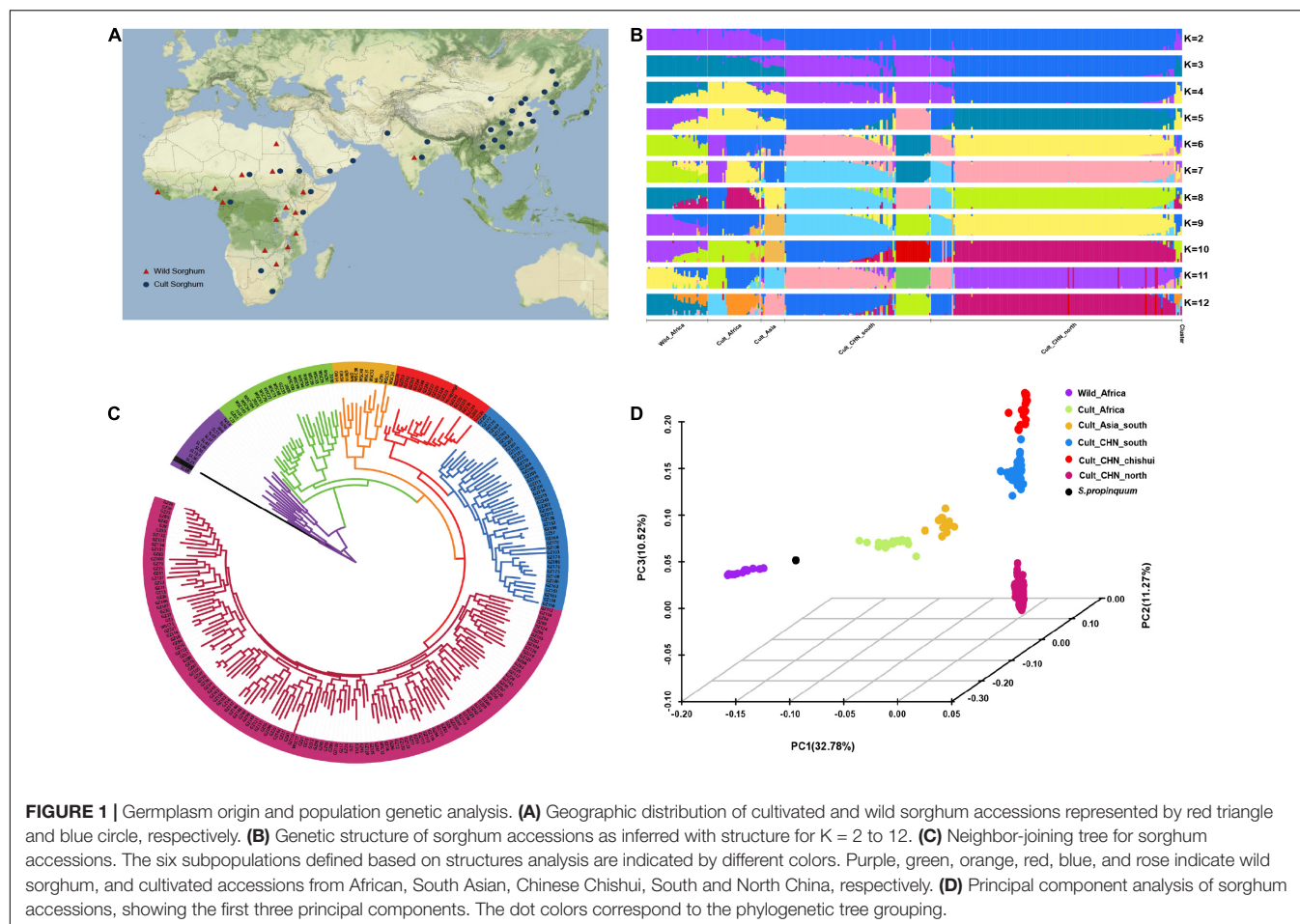
We considered hereafter a genotype to be unequivocally assigned to a population when its assignment probability was $\geq 95\%$ to one of the six genetic clusters at $K = 10$, which was the case for 71.5% of individuals ($N = 238$). It is important to set assignment thresholds in order to distinguish recent hybridization (or recent gene flow) from ancient gene flow, and the threshold we set appeared to be optimal (Groppe et al., 2021). Our maximum likelihood based phylogenetic dendrogram clearly distinguished six clusters (Figure 1C). The wild sorghum (purple) and African cultivated lines (green) split into two clades. The latter includes two forms of sorghums in north and south Africa, which was closely followed by South Asian sorghum (orange) and Chinese sorghum. Among Chinese accessions, sorghum from the Chishui (red) had a closer relationship with South Asian sorghum, while Northern sorghum (rose) was clearly separated from South Asian sorghum.

The above classification based on *fastSTRUCTURE* was also supported by principal component analysis. A total of 53.57% of the genetic variation was explained by the first principal component analysis (Figure 1D). The wild sorghum (purple), African cultivated sorghum (green), South Asian cultivated sorghum (orange) and Chinese accessions were well distinguished along the PC_1 axis, while the sorghum of Northern China (rose), Southern (blue) and Chishui (red) were obviously separated by PC_3 axis. South China sorghum were observed closer to South Asian sorghum than to the North China sorghum, which was supported by the phylogenetic relationship.

Genetic Diversity, Linkage Disequilibrium and Selective Sweep on Sorghum Collections

Genetic divergence among the inferred six genetic groups of sorghum was firstly estimated by the genetic distance (Fixation index values, F_{st}) across the genome (Figure 2A). The wild African sorghum had a lowest F_{st} (0.43) compared with the cultivated African sorghum, but the highest F_{st} (0.75) against the Chinese Chishui sorghum. However, the cultivated African sorghum showed lower F_{st} (0.48) with the Chinese sorghum in the south than that (0.53) of the north. These results were consistent with the phylogenetic and structure analyses.

We further implemented the mean nucleotide diversity (π) and Tajima's D to address the genetic diversity of the collections (Figures 2A,B). The π in African and Asian cultivated sorghum genomes were 1.14×10^{-3} and 1.03×10^{-3} , respectively. Low levels of π diversity were found in Chinese sorghum from the Northern, the Southern, and the Chishui sub-clusters (7.21×10^{-4} , 7.18×10^{-4} , and 5.52×10^{-4} , respectively). More



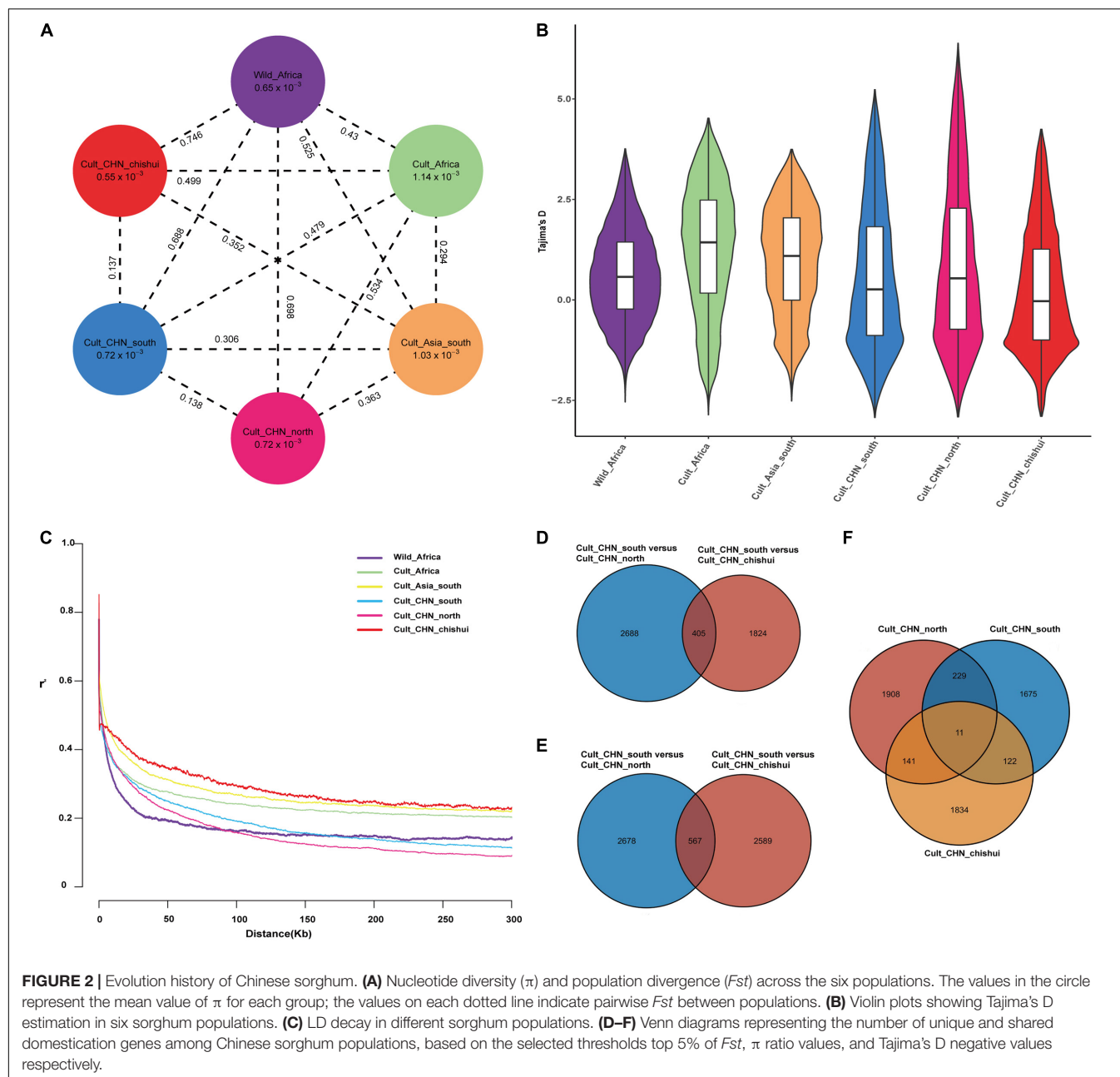
than half of Tajima's D values observed from Chishui sorghum genome were negative.

In our collections, the LD extended to background level at ~ 50 kb in wild African sorghum, and at ~ 100 , ~ 110 , and ~ 200 kb for African, South-Asian and Chinese cultivated sorghum, respectively (Figure 2C). The LD varied on different chromosomes and in different sorghum populations (Supplementary Figure 1). The wild African sorghum possessed rapid LD declines on both chromosome 6 and 10. The Southern sorghum appeared to have higher LD level than the Northern sorghum, and the Chishui sorghum showed the lowest LD decay on chromosome 5.

To identify the potential domestication signatures in Chinese sorghum, we summarized the selective sweeps within 10-kb genomic intervals and scored significant selection by top 5% outliers in F_{st} , π ratio and Tajima's D tests. Pairwise F_{st} statistics, respectively, revealed 3,093 and 2,229 candidate genes from the sweeps between Southern and Northern sorghum, and between Southern and Chishui sorghum, of which 405 candidate genes were shared (Figure 2D and Supplementary Figure 2). Meanwhile, candidate genes behind the selection signatures were identified as 3,245 by $\pi_{Southern}/\pi_{Northern}$ ratio and as 3,156 by $\pi_{Southern}/\pi_{Chishui}$ ratio, of which 567 candidate genes overlapped in the genomic regions (Figure 2E and

Supplementary Figure 2). Comparing the overlapping candidate genes between F_{st} and π ratio methods, 1,050 and 285 identical genes were found between Northern and Southern sorghums and Southern and Chishui sorghums, respectively. Additionally, the negative Tajima's D indicator was used to investigate the signals of the positive selections. We identified 2,289, 2,036, and 2,108 candidate genes for Southern, Northern and Chishui sorghum, respectively. Sorghum accessions from the Southern China shared 229 selective candidate genes against those from the Northern China and shared 122 candidate genes with Chishui sorghum. Totally, eleven candidate genes behind positive selection were in common among the three Chinese sorghum groups (Figure 2F and Supplementary Figure 3).

To understand putative functions of the strong selective sweep signals among Chinese populations, we investigated the annotations for the top 1% of the domestication genes identified by three methods. Only about 20% of these domesticated genes were matched to putative functions due to the lack of annotation. Overall, around half of the important genes are related to biotic and abiotic stress resistance, such as salt tolerance, cadmium tolerance, disease resistance, etc., while approximately 30% orthologous genes are associated with yield, quality, and plant type. The functions of important domestication genes did not differ significantly between populations based on F_{st}



and π ratio, whereas the functions of important domesticated genes were quite distinct in different groups based on Tajima's D metric. There were more key domestication genes related to panicle structure, tillering, height, and yield in southern and Chishui populations compared with those in northern population (Supplementary Table 3).

Impact of Selection for Traits Associated With Liquor-Brewing on Sorghum Genome

Since more than 30 processes are required to brew Maotai-flavor liquor in the Chishui River Basin in Southwest China,

the sorghum grains are required to have good cooking resistance. GWAS was thence performed to determine signals for the domestication traits-related brewing properties, based on surveying grain physical properties TGW, GL, GW, PT, and TT of the 244 Chinese landraces/cultivars in six environments from 2018 to 2020 (Supplementary Table 4). Totally, 66 significant SNP loci were identified on the 15 genome regions (Supplementary Table 5).

Two significant GWAS signals affecting both TGW and GW between Southern and Northern population during domestication were mapped in the selection sweeps associated with traits for liquor brewing were identified in this study (Supplementary Tables 3, 5). One important locus at the

position 47,459,472 bp on chromosome 9 fell into a LD block including several genes, of which the gene *Sobic.009G124200* (47,743,018 bp) is one of the strongest domestication genes based on *Fst* metric. This gene is an ortholog of the rice *SMALL ORGAN SIZE1* (*SMOS1*), which encodes an AP2-Type transcription factor acting as an auxin-dependent regulator for cell expansion during grain development (Aya et al., 2014). Another significant signal for TGW and GW was located on chromosome 10 at 11,199,795 bp position and underwent selection in both Southern and Chishui population during domestication based on Tajima's *D* metric (Figures 3A,B). One gene *Sobic.010G111200* (11,197,868 bp) fell within LD block of significant loci (Figure 3C) and is orthologous to rice *OsGASR7/GW6* that encodes a GA-regulated GAST family protein and positively regulates grain width and weight (Shi et al., 2020). As expected, Chishui and Southern sorghums have smaller grains than Northern sorghum (Figure 3D). The mRNA expression of *Sobic.010G111200* in developing panicles of less than 5-cm-long was significantly higher in cultivar LTR108 (larger grain) than in cultivar 654 (smaller grain), while mRNA levels in LTR108 dropped dramatically with panicle development and was even lower than the level of 654 after heading (Figure 3E).

We also identified two GWAS signals for TT and PT (Supplementary Table 5). A GWAS signal associated with TT was detected on chromosome 2 at 8,300,524 bp position, which is clearly distinct from the previously reported *B2* locus (60.54–60.59 Mb) that control the presence of testa (Dufour et al., 1997; Rami et al., 1998; Mace et al., 2019). In addition, a significant signal was identified for pericarp thickness on chromosome 2 (61.99 Mb) that is close to the previously reported *Z* gene locus (57.03–59.10 Mb), which is related to the thickness of pericarp and pearly pericarp (Tao et al., 1998; Boivin et al., 1999; Mace et al., 2019). The signal was located in a large genomic region (140 kb, 61.96–62.10 Mb) of the strongly selective sweep during domestication in Chishui population based on Tajima's *D* metric (Figures 4A,B). The phenotypic distribution indicated that the pericarp of Chishui and southern accession was thinner than that of northern accession (Figures 4D,E). The significant SNP loci fell within a high LD block including two genes (*Sobic.002G228600* and *Sobic.002G228700*) (Figure 4C). By comparing the sequence mutations of *Sobic.002G228600* gene between extreme phenotypic varieties, we detected an insertion and deletion of "CATATTACAATCC" in the 5'-UTR in accessions with thin and thick pericarp, respectively (Figure 4F and Supplementary Figure 4).

Relationship Between Baijiu-Brewing Process and the Physical Properties of Grains

Chishui sorghum has consistently thinner pericarp (exocarp and pericarp) than the other two Chinese populations (Figure 3E). Correlation analysis revealed that sorghum pericarp thickness showed significant positive association with TGW, GL and GW in six environments, with the average Pearson coefficient (*R* value) of 0.32, 0.24 and 0.26 ($P < 0.01$), respectively

(Supplementary Table 6). The cooking experiment was performed to investigate the relationship between resistance to cooking with grain size and thickness of pericarp and testa. For glutinous sorghum, grain with thinner pericarp and testa generally had lower the rate of expansion and crack and presented better resistance to cooking (Table 1).

DISCUSSION

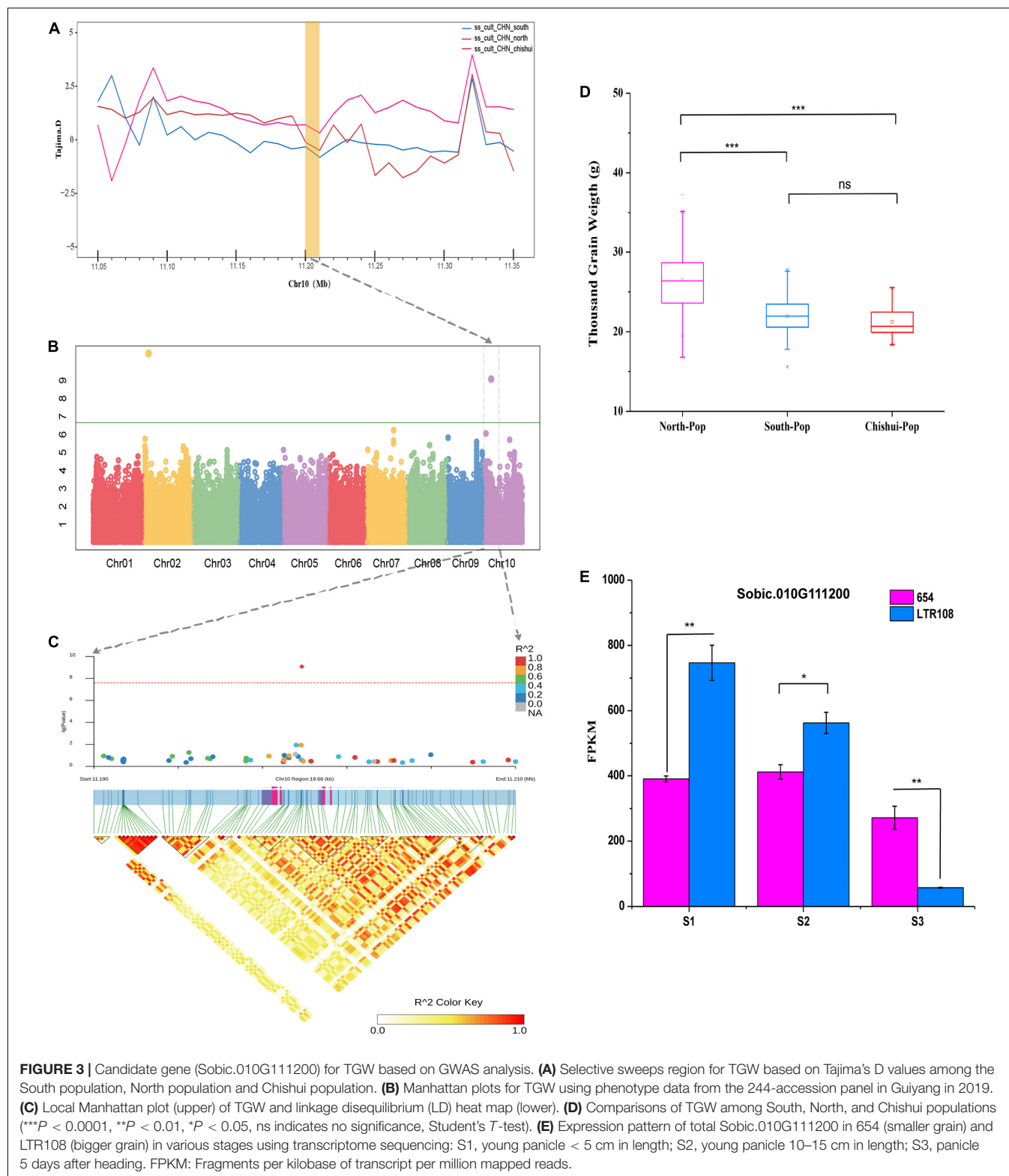
The Origin of Chinese Sorghum

A comprehensive understanding of the origin and domestication history of crops will help us better manage and use it. In this study, we conducted substantial investigations on population genomics and evolutionary history of Chinese sorghum based on resequencing of 333 diverse accessions. Our population genetics analyses revealed that sorghum accessions from South China were more closely related to wild relatives and landraces/varieties from African and South Asian. Similar patterns were observed in Chinese castor originated in Africa (Fan et al., 2019) and Chinese cotton originated in India (Du et al., 2018), Southern varieties are more closely related to wild relatives than Northern varieties. As the local residents in the southwestern border of China have historically maintained close contact with the people of the Indochina Peninsula and the Indian subcontinent (You, 1989). This is also the origin of the ancient Chinese name of sorghum, "ShuShu," which means millet-like crop grown in the Kingdom of Shu, located in the southwestern region of China (Zhao, 2019a). We hence believe that sorghum may have been introduced from India to the southwestern region of China and subsequently spread into middle and northern China, supporting the hypothesis that sorghum entered China through the Southwest Silk Road (Zhao et al., 2019). Due to the limited sampling size (38 African and 21 Asian cultivated accessions) used in this study, this conclusion is considered tentative and needs further validation with larger sample sizes.

Linkage Disequilibrium and Selective Sweep on Sorghum Collections

Sorghum, as an often cross-pollinated crop, has a lower linkage disequilibrium (LD) decay rate than self-pollinated rice (120 kb and 160 kb for indica and japonica, respectively) (Xu et al., 2011), and much higher than open-pollinated maize (~30 kb) (Hufford et al., 2012). The LD from wild sorghum to Chinese cultivated accession extended to background level from ~50 to ~200 kb in our collection. This is consistent with LD estimates in sorghum by using GBS and WGS techniques (Mace et al., 2013; Morris et al., 2013). In addition, chromosomes with the slowest decay rate were different in different Chinese sorghum populations, and similar phenomenon was also observed in bread wheat (Hao et al., 2020). The possible reasons are that local agroecological environments and human preference, or some genes for important agronomic traits clustering in the same segment, which causes each chromosome to experience different selection pressures during domestication in the different populations.

Our research also revealed that more genes have been subject to domestication between southern and northern populations



than southern and Chishui populations, suggesting that the number of selective sweep signals varies between different populations. Likewise, around half of the important genes is related to biotic and abiotic stress resistance, while approximately

30% orthologous genes are associated with yield, quality, and plant height. This confirms that crop domestication is the result of a synergistic effect of natural and artificial selection (Smith et al., 2019).

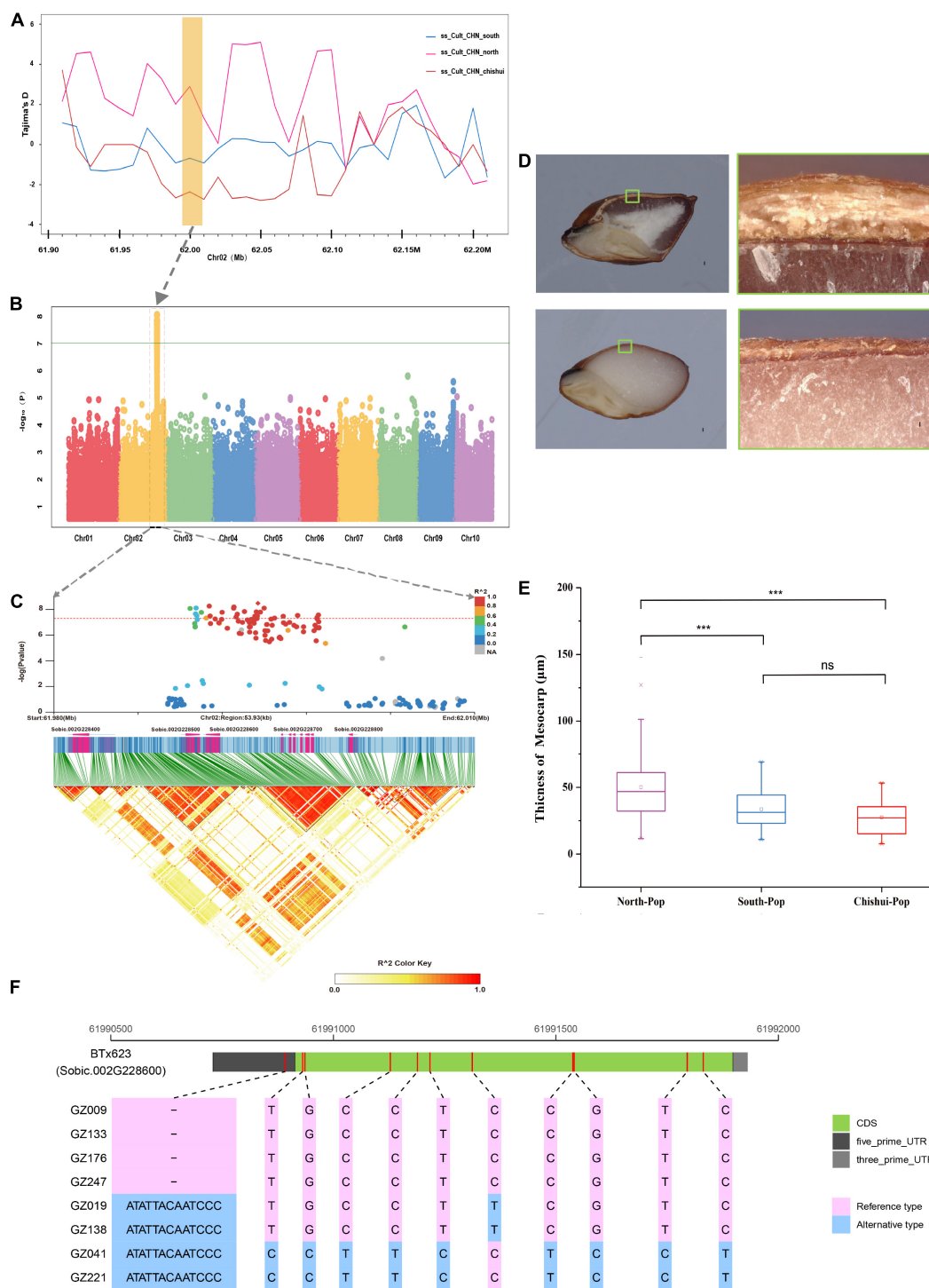


FIGURE 4 | Candidate gene for pericarp thickness based on genome-wide association study analysis. **(A)** The selective sweep analysis of pericarp thickness genes based on Tajima's D values among the South population, North population and Chishui population. **(B)** Manhattan plots for pericarp thickness using phenotype data from the 244-accession panel in Ledong in 2020. **(C)** Local Manhattan plot (upper) of pericarp thickness and linkage disequilibrium (LD) heat map (lower). **(D)** Longitudinal sections of intact seeds from thin pericarp accession in upper right image and thick pericarp sorghum accession in lower right image. The upper and lower images on the left present close-ups of the pericarp under the microscope, respectively corresponding to the small boxes in the right images. **(E)** Comparisons of pericarp thickness among South population, North population, and Chishui population ($***P < 0.0001$, ns indicates no significance, Student's *T*-test). **(F)** The putative candidate gene model of pericarp thickness for eight extreme phenotypes. The red bar marks the mutant position between reference type and alternative type. INDEL in 5'-UTRs with pink indicates the reference genotype (thinner pericarp), while that with blue indicates the alternative genotype (thicker pericarp).

TABLE 1 | Results of cooking experiment for glutinous sorghum.

Name	Pericarp thickness	Testa thickness	Total thickness	Thousand grain weight	Expansion rate \pm SD	Crack rate \pm SD
GZ243	27.82	13.03	40.85	16.57	115.18% \pm 9.88%	83.00% \pm 6.67%
GZ060	36.49	14.25	50.74	20.21	114.78% \pm 4.94%	81.00% \pm 3.33%
GZ241	45.89	12.64	58.52	22.90	111.43% \pm 5.37%	90.00% \pm 0%
GZ136	48.20	17.37	65.57	23.81	91.67% \pm 10.02%	77.50% \pm 2.33%
GZ162	50.05	18.87	68.91	25.77	138.89% \pm 7.84%	96.50% \pm 0%
GZ093	50.20	4.45	54.64	29.13	117.1% \pm 7.26%	73.50% \pm 4.43%
GZ150	56.84	29.28	86.12	30.62	92.11% \pm 2.34%	9.00% \pm 0.67%
GZ210	72.35	7.75	80.10	19.18	111.43% \pm 2.61%	81.50% \pm 1.67%
GZ234	76.05	16.47	92.52	21.86	102% \pm 2.61%	89.50% \pm 2.33%
GZ202	84.60	23.13	107.73	16.23	126.87% \pm 10.75%	96.50% \pm 2.33%
GZ187	88.81	17.53	106.34	22.29	169.44% \pm 5.74%	99.00% \pm 0.67%
GZ192	88.90	14.59	103.49	17.46	169.09% \pm 9.30%	86.50% \pm 4.33%
GZ104	97.01	16.44	113.45	28.63	146.02% \pm 1.16%	93.00% \pm 0%
GZ220	99.47	16.32	115.78	26.26	133.33% \pm 0%	98.00% \pm 3.33%
GZ061	135.85	34.45	170.30	21.81	190.71% \pm 10.02%	100.00% \pm 0%
GZ123	157.34	0.00	157.34	26.56	151.08% \pm 5.16%	94.50% \pm 1.00%

Domestication of Sorghum Grain Physical Properties Related to Liquor Brewing Process

The Southwest region in China, especially in the Chishui River Basin between Guizhou and Sichuan province, is the birthplace and main producing area of Maotai-aroma liquor. It is generally believed that the glutinous sorghum with features of small grain and thick hull can withstand 30 processes and several rounds of steaming and boiling in brewing Maoutai's. Phenotype data showed Chishui and Southern sorghums have smaller grains than Northern sorghum, and the GWAS identified two important domestication genes *Sobic.009G124200* and *Sobic.010G111200* for TGW and GW. The two genes are an ortholog of the *SMALL ORGAN SIZE1 (SMOS1)* and *OsGASR7/GW6* controlling grain size in rice (Aya et al., 2014; Shi et al., 2020). The varieties possessing longer and looser panicles with pendulous branches are more suitable for the hot and humid climate in southern China in summer due to their resistance to mold and insects at the stage of maturity, whereas sorghum panicle length is negatively correlated with grain size (Tao et al., 2021), indicating that natural selection for panicle length genes in southern and Chishui populations is likely to cause selective sweep of genes controlling grain size.

Grain pericarp thickness failed to catch research's attention, probably because they have no obvious relationship with traits such as yield, quality, and resistance to biotic and abiotic stresses. The *Z* gene has not been extensively studied since it was firstly reported in 1930s (Ayyangar and Ayyar, 1938), which is often used as a morphological marker to construct genetic maps in sorghum (Tao et al., 1998; Boivin et al., 1999; Mace et al., 2019). Varieties with light-colored grains are easy to determine whether pericarp is pearly, but those with dark-colored grains are difficult to determine whether pericarp is pearly, which might lead to an inaccurate genetic mapping of *Z* gene on chromosome. Based on the observation under a digital microscope, our GWAS analysis identified an important loci of pericarp thickness that is ~ 2 Mb away from the previously reported *Z* gene locus (Mace et al., 2019). The possible candidate

gene *Sobic.002G228600* is an ortholog of *GRMZM5G898880* involving in striga resistance in maize (Adewale et al., 2020), and a QTL for striga resistance was reported in nearby areas (60.58–60.80 Mb) in a previous sorghum study (Haussmann et al., 2004). *Sobic.002G228600* gene showed an insertion and deletion of "CATATTACAATCC" in the 5'-UTR in accessions with thin and thick pericarp. 5'-UTR contains key elements of translational regulation, such as structural motifs and upstream open reading frames (uORFs). By controlling the selection of translation initiation sites (TISs), many sequence elements in 5'-UTR contribute to mRNA translatability (Hinnebusch et al., 2016). For example, the significant InDels in the 5'-UTR alter the expression of *BRB* (Big Root Biomass) in sesame (Dossa et al., 2021), and an 8-bp InDel in the 5'-UTR of *SlbHLH59* regulates ascorbate biosynthesis in tomato (Ye et al., 2019).

We found that *Z* gene is located in a ~ 140 kb (61.96–62.10 Mb) regions on chromosome 2 that underwent the strong artificial selection in Chishui population. In addition to grain pericarp thickness, this significant region is likely to be related to grain size, panicle shape, stem morphology and disease resistance. Several genes in the vicinity of *Z* gene locus, such as *Sobic.002G228800*, *Sobic.002G228900*, *Sobic.002G229000*, and *Sobic.002G229100*, are homologous to the genes controlling panicle architecture and seed size in Arabidopsis, rice, and maize. The gene *Sobic.002G228700* encodes an ortholog of *PATATIN-RELATED PHOSPHOLIPASE A (DEP3)*, which affects the architecture of inflorescence in rice (Qiao et al., 2011), as well as the cellulose content and cell elongation in Arabidopsis (Li M. et al., 2011). The gene *Sobic.002G228800* encodes an ortholog of *C-TERMINALLY ENCODED PEPTIDE (OsCEP6.1)* in rice, which regulates the development of panicle type and grain size by changing cell size without altering cell number (Sui et al., 2016). Two genes (*Sobic.002G229000* and *Sobic.002G229100*) encode not only an ortholog of a *PUTATIVE SERINE CARBOXYPEPTIDASE (GS5)* as a positive regulator of rice grain size (Li Y. et al., 2011), but also an ortholog of *SERINE CARBOXYPEPTIDASE-LIKE 40 (GRMZM2G072240)* involving in formation of the number of hull layers in maize (Cui et al., 2018). Several QTLs for panicle shape, thousand grain weight,

and grain yield were consistently mapped in the same genomic regions as in the previous studies in sorghum (Mantilla Perez et al., 2014; Zhou et al., 2019). Moreover, many QTLs related to disease resistance and morphology were also reported in this genomic region, such as shoot fly resistance (Aruna et al., 2011), ergot resistance (Parh et al., 2008), downy mildew resistance (Ahn et al., 2019), rust resistance (Wang et al., 2014), plant height (Shiringani et al., 2010; Mantilla Perez et al., 2014), tillering number (Alam et al., 2014; Kong et al., 2015), polyphenol content (Rhodes et al., 2017). This is in line with the main traits of Chishui sorghum varieties, which possess smaller grain size, looser panicle, better disease resistance and higher plant height, to adapt to local agroecological environments, such as hot and humid summer, serious pests and diseases. Compared with the thickness of pericarp, it is easier for local farmers to select new varieties based on traits such as grain size, panicle shape, and disease resistance during the domestication and improvement of sorghum. As a result, the strong artificial selection of grain size, panicle type or resistance to biotic stress is most likely to cause the selective sweep, which led to a genetic hitchhiking of nearby *Z* gene conferring pericarp thickness and a large reduction of the genetic diversity in the 140 kb regions on chromosome 2.

Relationship Between Baijiu-Brewing Process and the Physical Properties of Grains

Traditionally, Moutai-aroma liquor brewing has always preferred glutinous sorghum varieties with smaller size and thicker pericarp, which were thought to meet the process technology of repeated steaming (Chen and Ji, 2006). Contrary to the conventional belief, our results revealed that the increase in pericarp thickness results in larger grain size, and that pericarp thickness rather than grain size is associated with the rate of expansion and crack in waxy sorghum, meaning that smaller-size grains have thinner pericarp, while grain with thinner pericarp generally exhibit better resistance to cooking. The similar results were reported in the effect of grain physical properties on sorghum processing. A thick pericarp absorbs more water and is more easily detached during the dehulling than a thin pericarp (Scheuring et al., 1983). Our findings also suggest that selection for *Z* gene should be avoided in sorghum high-yield breeding program, because the high-yielding varieties caused by thick pericarp have lower rate of dehulled grains and flour yield during milling (Lu, 1999).

In summary, population structure analysis revealed that 333 sorghum accessions clustered into six groups of wild relatives, cultivated sorghum in Africa, Southern Asia, Southern China, Northern China, and Chishui in Southwest China, supporting the hypothesis that sorghum entered China through the Southwest Silk Road. We discovered a significant selective sweep that harbors several genes for pericarp thickness, grain size, panicle

type or biotic stress tolerance, which resulted in a large reduction of the genetic diversity in the 140 kb regions on chromosome 2. Based on the cooking experiment, we put forward a more scientific standard for sorghum breeding that is suitable for liquor brewing technology in China. Our results provided useful information for the breeding and improvement not only for liquor-sorghum but also for other breeding programs in China.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: China National GeneBank (CNCB); CNP0002968.

AUTHOR CONTRIBUTIONS

LZ and GZ designed and managed the research and wrote the manuscript. YD, JX, XG, KL, ZF, NC, BC, LBZ, MR, and ZX collected the plant materials, investigated traits, and performed the experiments. YD, XL, and ZB analyzed whole-genome sequence data and RNA-seq data. ZX and YT revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (32160459), the National Key R&D Program of China (2018YFD1000706/2018YFD10007011), Guizhou Natural Science Foundation of China [QKHJC(2020)1Y103], Zhejiang Major Scientific and Technological Project of Agricultural (Upland crop) Breeding (2021C02064-6).

ACKNOWLEDGMENTS

We thank Liu Shuo from Liaoning Institute of Pomology, China, for the enthusiasm and support in bio-information analysis and revision on the manuscript. We also thank Li Guangwei from Henan Agricultural University, China, for the comments on the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.923734/full#supplementary-material>

REFERENCES

- Adewale, S., Badu-Apraku, B., Akinwale, R., Paterne, A., Gedil, M., and Garcia-Oliveira, A. (2020). Genome-wide association study of *Striga* resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol.* 20:203. doi: 10.1186/s12870-020-02360-0
- Ahn, E., Hu, Z., Perumal, R., Prom, L., Odvody, G., Upadhyaya, H., et al. (2019). Genome wide association analysis of sorghum mini core lines regarding

- anthracnose, downy mildew, and head smut. *PLoS One* 14:e0216671. doi: 10.1371/journal.pone.0216671
- Alam, M., Mace, E., van Oosterom, E., Cruickshank, A., Hunt, C., Hammer, G., et al. (2014). QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. *Theoret. Appl. Genet.* 127, 2253–2266. doi: 10.1007/s00122-014-2377-9
- Aruna, C., Bhagwat, V., Madhusudhana, R., Sharma, V., Hussain, T., Ghorade, R., et al. (2011). Identification and validation of genomic regions that affect shoot fly resistance in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoret. Appl. Genet.* 122, 1617–1630. doi: 10.1007/s00122-011-1559-y
- Awika, J., and Rooney, L. (2004). Sorghum phytochemicals and their potential impact on human health. *Phytochemistry* 65, 1199–1221. doi: 10.1016/j.phytochem.2004.04.001
- Aya, K., Hobo, T., Sato-Izawa, K., Ueguchi-Tanaka, M., Kitano, H., and Matsuoka, M. (2014). A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. *Plant Cell Physiol.* 55, 897–912. doi: 10.1093/pcp/pcu023
- Ayyangar, G., and Ayyar, M. (1938). Linkage between a panicle factor and the pearly-chalky mesocarp factor (Zz) in sorghum. *Proc. Ind. Acad. Sci. Sect. B.* 8, 100–107. doi: 10.1007/BF03048478
- Boivin, K., Deu, M., Rami, J. F., Trouche, G., and Hamon, P. (1999). Towards a saturated sorghum map using RFLP and AFLP markers. *Theoret. Appl. Genet.* 98, 320–328. doi: 10.1007/s001220051076
- Boivin, N., and Fuller, D. Q. (2009). Shell middens, ships and seeds: exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula. *J. World Prehist.* 22, 113–180. doi: 10.1007/s10963-009-9018-2
- Chen, B. R., Wang, C. Y., Wang, G. P., Zhu, Z. X., Xu, N., Shi, G. S., et al. (2019). Genome-wide association study for starch content and constitution in sorghum (*Sorghum bicolor* (L.) Moench). *J. Integr. Agricult.* 18, 2446–2456. doi: 10.1016/S2095-3119(19)62631-6
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, X. X., and Ji, K. L. (2006). General Introduction to the individualities of Maotai Liquor. *Liquor-Mak. Sci. Technol.* 140:5.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Cui, Z., Xia, A., Zhang, A., Luo, J., Yang, X., Zhang, L., et al. (2018). Linkage mapping combined with association analysis reveals QTL and candidate genes for three husk traits in maize. *Theoret. Appl. Genet.* 131, 2131–2144. doi: 10.1007/s00122-018-3142-2
- Dahlberg, J. A., and Wasylukowa, K. (1996). Image and statistical analyses of early sorghum remains (8000 b.p.) from the Nabta Playa archaeological site in the Western Desert, southern Egypt. *Veget. Hist. Archaeobot.* 5:6.
- Danecek, P., Auton, A., Abecasis, G., Albers, C., Banks, E., DePristo, M., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Doggett, H. (1988). Sorghum. *Econ. Bot.* 1, 355–371.
- Dong, S., He, W., Ji, J., Zhang, C., Guo, Y., and Yang, T. (2021). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* 22:227. doi: 10.1093/bib/bbaa227
- Dossa, K., Zhou, R., Li, D., Liu, A., Qin, L., Mmadi, M. A., et al. (2021). A novel motif in the 5'-UTR of an orphan gene 'Big Root Biomass' modulates root biomass in sesame. *Plant Biotechnol. J.* 19, 1065–1079. doi: 10.1111/pbi.13531
- Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nature genetics* 50, 796–802. doi: 10.1038/s41588-018-0116-x
- Dufour, P., Deu, M., Grivet, L., D'Hont, A., Paulet, F., Bouet, A., et al. (1997). Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. *Theoret. Appl. Genet.* 94, 409–418. doi: 10.1007/s001220050430
- Fan, W., Lu, J., Pan, C., Tan, M., Lin, Q., Liu, W., et al. (2019). Sequencing of Chinese castor lines reveals genetic signatures of selection and yield-associated loci. *Nat. Comm.* 10:3418. doi: 10.1038/s41467-019-11228-3
- Fuller, D. Q. (2003). African crops in prehistoric South Asia: A critical review. In: Neumann, K., Butler, A., Kahlheber, S. (Eds.), *Food, Fuel and Fields: Progress in Africa Archaeobotany. Africa Praehist.* 15:32.
- Goodstein, D., Shu, S., Howson, R., Neupane, R., Hayes, R., Fazo, J., et al. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Groppi, A., Liu, S., Cornille, A., Decroocq, S., Bui, Q., Tricon, D., et al. (2021). Population genomics of apricots unravels domestication history and adaptive events. *Nat. Comm.* 12:3956. doi: 10.1038/s41467-021-24283-6
- Han, M. L. (2012). *Historical Agricultural Geography of China*. Beijing: Peking University Press.
- Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y., et al. (2020). Resequencing of 145 Landmark Cultivars Reveals Asymmetric Sub-genome Selection and Strong Founder Genotype Effects on Wheat Breeding in China. *Mole. Plant* 13, 1733–1751. doi: 10.1016/j.molp.2020.09.001
- Hausmann, B., Hess, D., Omany, G., Folkertsma, R., Reddy, B., Kayentao, M., et al. (2004). Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *Theoret. Appl. Genet.* 109, 1005–1016. doi: 10.1007/s00122-004-1706-9
- Hinnebusch, A., Ivanov, I., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416. doi: 10.1126/science.aad9868
- Hufford, M., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J., Cartwright, R., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309
- Joseph, N. (2000). *Science and civilisation in China*. Cambridge, MA: Cambridge University Press.
- Kong, W., Kim, C., Goff, V., Zhang, D., and Paterson, A. (2015). Genetic analysis of rhizomatousness and its relationship with vegetative branching of recombinant inbred lines of *Sorghum bicolor* × *S. propinquum*. *Am. J. Bot.* 102, 718–724. doi: 10.3732/ajb.1500035
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, M., Bahn, S., Guo, L., Musgrave, W., Berg, H., Welti, R., et al. (2011). Patatin-related phospholipase pPLAIIIβ-induced changes in lipid metabolism alter cellulose content and cell elongation in Arabidopsis. *Plant Cell* 23, 1107–1123. doi: 10.1105/tpc.110.081240
- Li, R. Y., Zhang, H., Zhou, X. C., Guan, Y. N., Yao, F. X., Song, G. A., et al. (2010). Genetic diversity in Chinese sorghum landraces revealed by chloroplast simple sequence repeats. *Genet. Resour. Crop Evol.* 57, 1–15. doi: 10.1007/s10722-009-9446-y
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43, 1266–1269. doi: 10.1038/ng.977
- Liu, M. K., Tang, Y. M., Ren, D. Q., Yao, W. C., Ni, B., and Lei, G. D. (2012). Brewing property comparison between several species of sorghum seed. *China Brew.* 31:4.
- Lu, Q. S. (1999). *Sorghum*. Beijing: China Agriculture Press.
- Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., et al. (2019). The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theoret. Appl. Genet.* 132, 751–766. doi: 10.1007/s00122-018-3212-5
- Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentis, P. J., Bian, L., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Comm.* 4:2320. doi: 10.1038/ncomms3320
- Mann, J. A., Kimber, C. T., and Miller, F. R. (1983). The origin and early cultivation of sorghums in Africa. *Bulletin - Texas Agricult. Exp. Stat.* 1983:21. doi: 10.1007/s10437-018-9314-2
- Mantilla Perez, M., Zhao, J., Yin, Y., Hu, J., and Salas Fernandez, M. (2014). Association mapping of brassinosteroid candidate genes and plant architecture in a diverse panel of Sorghum bicolor. *Theoret. Appl. Genet.* 127, 2645–2662.
- McCormick, R., Truong, S., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2018). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tpj.13781

- 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
- Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci.* 110, 453–458. doi: 10.1073/pnas.1215985110
- Parh, D., Jordan, D., Aitken, E., Mace, E., Jun-ai, P., McIntyre, C., et al. (2008). QTL analysis of ergot resistance in sorghum. *Theoret. Appl. Genet.* 117, 369–382. doi: 10.1007/s00122-008-0781-8
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723
- Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qiao, Y., Piao, R., Shi, J., Lee, S., Jiang, W., Kim, B., et al. (2011). Fine mapping and candidate gene analysis of dense and erect panicle 3, *DEP3*, which confers high grain yield in rice (*Oryza sativa* L.). *Theoret. Appl. Genet.* 122, 1439–1449. doi: 10.1007/s00122-011-1543-6
- Raj, A., Stephens, M., and Pritchard, J. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350
- Rami, J. F., Dufour, P., Trouche, G., Fliedel, G., Mestres, C., Davrieux, F., et al. (1998). Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theoret. Appl. Genet.* 97, 605–616. doi: 10.1007/s001220050936
- Rhodes, D., Gadgil, P., Perumal, R., Tesso, T., and Herald, T. J. (2017). Natural Variation and Genome-Wide Association Study of Antioxidants in a Diverse Sorghum Collection. *Cereal Chem.* 94, 190–198. doi: 10.1094/CCHEM-03-16-0075-R
- Scheuring, J. F., Sidibe, S., Rooney, L. W., and Earp, C. F. (1983). Sorghum pericarp thickness and its relation to decortication in a wooden mortar and pestle. *Cer. Chem.* 60, 86–89.
- Shi, C., Dong, N., Guo, T., Ye, W., Shan, J., and Lin, H. (2020). A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. *Plant J.* 103, 1174–1188. doi: 10.1111/tpj.14793
- Shinoda. (1958). *Baijiu – The introduction of sorghum Taiwan*. Elite Publishing Co., Ltd.
- Shiringani, A., Frisch, M., and Friedt, W. (2010). Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. *Theoret. Appl. Genet.* 121, 323–336. doi: 10.1007/s00122-010-1312-y
- Smith, O., Nicholson, W., Kistler, L., Mace, E., Clapham, A., Rose, P., et al. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. *Nat. Plants* 5, 369–379. doi: 10.1038/s41477-019-0397-9
- Sui, Z., Wang, T., Li, H., Zhang, M., Li, Y., Xu, R., et al. (2016). Overexpression of Peptide-Encoding *OsCEP6.1* Results in Pleiotropic Effects on Growth in Rice (*O. sativa*). *Front. Plant Sci.* 7:228. doi: 10.3389/fpls.2016.00228
- Sun, B. G. (2019). *Chinese National Alcohols*. Beijing: China Chemical Industry Press.
- Tao, Y., Trusov, Y., Zhao, X., Wang, X., Cruickshank, A., Hunt, C., et al. (2021). Manipulating assimilate availability provides insight into the genes controlling grain size in sorghum. *Plant J.* 108, 231–243. doi: 10.1111/tpj.15437
- Tao, Y. Z., Jordan, D. R., Henzell, R. G., and McIntyre, C. L. (1998). Construction of a genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps. *Austr. J. Agric. Res.* 49, 729–736. doi: 10.1071/A97112
- Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3
- Wang, X., Mace, E., Hunt, C., Cruickshank, A., Henzell, R., Parkes, H., et al. (2014). Two distinct classes of QTL determine rust resistance in sorghum. *BMC Plant Biol.* 14:366. doi: 10.1186/s12870-014-0366-4
- Winchell, F., Stevens, C. J., Murphy, C., Champion, L., and Fuller, D. Q. (2017). Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan Spikelet Morphology from Ceramic Impressions of the Butana Group. *Curr. Anthropol.* 58, 673–683.
- Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., et al. (2018). Diversification and independent domestication of Asian and European pears. *Gen. Biol.* 19:77. doi: 10.1186/s13059-018-1452-y
- Xu, W., Wu, D., Yang, T., Sun, C., Wang, Z., Han, B., et al. (2021). Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean. *Gen. Biol.* 22:113. doi: 10.1186/s13059-021-02333-y
- Xu, X., Liu, X., Ge, S., Jensen, J., Hu, F., Li, X., et al. (2011). 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
- Ye, J., Li, W., Ai, G., Li, C., Liu, G., Chen, W., et al. (2019). Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis via D-mannose/L-galactose pathway in tomato. *PLoS Genet.* 2019:15. doi: 10.1371/journal.pgen.1008149
- You, X. L. (1989). The time and origin of the introduction of corn into China and Asia. *Ancient Mod. Agricult.* 2:9.
- Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., et al. (2018). Origin and evolution of qingke barley in Tibet. *Nat. Comm.* 9:5433. doi: 10.1038/s41467-018-07920-5
- Zhang, C., Dong, S., Xu, J., He, W., and Yang, T. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875
- Zhang, H., Wang, J.-C., Wang, D.-J., Yao, F.-X., Xu, J.-F., Song, G.-A., et al. (2011). Assessment of genetic diversity in Chinese sorghum landraces using SSR markers as compared with foreign accessions. *Crop J.* 37, 224–234. doi: 10.3724/SP.J.1006.2011.00224
- Zhang, L., Leng, C., Luo, H., Wu, X., Liu, Z., Zhang, Y., et al. (2018). DrySweet Sorghum Originated through Selection of, a Plant-Specific NAC Transcription Factor Gene. *Plant Cell* 30, 2286–2307. doi: 10.1105/tpc.18.00313
- Zhao, L. J. (2019b). Study on Ancient Names of Sorghum in the Pre-Qin and Han Dynasties. *Agricult. Archaeol.* 2019, 36–45.
- Zhao, L. J. (2019a). Discuss on the Time, Path of Sorghum's Introduction into China and its Preliminary Popularization. *Agricult. Hist. Chin.* 2019:1.
- Zhao, Y. P., Fan, G., Yin, P. P., Sun, S., Li, N., Hong, X., et al. (2019). Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil. *Nat. Comm.* 10:4201. doi: 10.1038/s41467-019-12133-5
- Zhou, Y., Srinivasan, S., Mirnezami, S., Kusmec, A., Fu, Q., Attigala, L., et al. (2019). Semiautomated Feature Extraction from RGB Images for Sorghum Panicle Architecture GWAS. *Plant Physiol.* 179, 24–37. doi: 10.1104/pp.18.00974
- Zou, G., Zhai, G., Yan, S., Li, S., Zhou, L., Ding, Y., et al. (2020). Sorghum qTGW1a encodes a G-protein subunit and acts as a negative regulator of grain size. *J. Exp. Bot.* 71, 5389–5401. doi: 10.1093/jxb/eraa277

Conflict of Interest: ZB was employed by Shanghai OE Biotech Co., Ltd., China.

The remaining 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Ding, Xu, Gao, Cao, Li, Feng, Cheng, Zhou, Ren, Lu, Bao, Tao, Xin and Zou. 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(s) 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.



Artificial Selection Trend of Wheat Varieties Released in Huang-Huai-Hai Region in China Evaluated Using DUS Testing Characteristics

Liyuan Wang, Yongsheng Zheng, Lili Duan, Mumu Wang, Hui Wang, Hua Li, Ruyu Li* and Han Zhang*

Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China

OPEN ACCESS

Edited by:

Xia Xin,
Institute of Crop Sciences
(CAAS), China

Reviewed by:

Quan Xie,
Nanjing Agricultural University, China
Weihua Qiao,
Institute of Crop Sciences
(CAAS), China

*Correspondence:

Ruyu Li
li_ruyu@sina.com
Han Zhang
peony-han@163.com

Specialty section:

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

Received: 17 March 2022

Accepted: 16 May 2022

Published: 10 June 2022

Citation:

Wang L, Zheng Y, Duan L, Wang M,
Wang H, Li H, Li R and Zhang H
(2022) Artificial Selection Trend of
Wheat Varieties Released in
Huang-Huai-Hai Region in China
Evaluated Using DUS Testing
Characteristics.
Front. Plant Sci. 13:898102.
doi: 10.3389/fpls.2022.898102

Wheat has been widely cultivated all over the world. In China, the number of approved wheat varieties has steadily grown since 2010, with the most notable trend in the Huang-Huai-Hai region. Distinctiveness, uniformity, and stability (DUS) are the prerequisites for a new wheat variety to obtain a release permit. Yet, few reports are available on DUS testing characteristics of released wheat varieties. Here, 32 DUS testing characteristics of 195 wheat varieties released in the Huang-Huai-Hai region were investigated to study their artificial selection trend. The results showed that the means, ranges, and coefficients of variation for eight measured characteristics varied greatly, among which the number of sterile spikelets had the largest variation coefficient of all three wheat-growing areas in the Huang-Huai-Hai region. The difference in plant height between the three wheat-growing areas was the most significant. The mean plant height in the northern winter wheat area was the largest, while that in south Huanghuai was the smallest. The released varieties of the three wheat-growing areas in the region had similar artificial selection trends in some characteristics. For instance, flag leaf length and flag leaf width, grain number per ear, and grain volume weight showed an overall upward trend, while the plant height gradually decreased. The clustering results based on DUS testing characteristics showed that artificial selection of characteristics was consistent with ecological adaptation and breeding process as well as pedigree sources. Our findings indicated that with the current breeding objectives, the selection of some non-economic characteristics of wheat varieties, such as awn color, stem color, and glume color, seemed to be able to enrich the genetic diversity of varieties in the Huang-Huai-Hai region. These results could provide guidance for subsequent wheat breeding and production in this region, screening similar varieties, and determining the distinctness of applied varieties in DUS testing.

Keywords: wheat varieties, DUS testing characteristics, Huang-Huai-Hai region, artificial selection trend, variation diversity

INTRODUCTION

Wheat is widely cultivated in the world, and it is also one of the three major grain crops in China. According to the National Bureau of Statistics, the sown area and output of wheat accounted for 20.44 and 20.13% of the total crop sown area and output of grain crops in China, respectively (2019). In China, the Huang-Huai-Hai region is the main wheat-growing area, accounting for 57.54 and 63.10% of the total wheat sown area and output, respectively. The Huang-Huai-Hai region is a large eco-geographical zone of about 469,500 square kilometers, roughly covering areas of Beijing, Tianjin, Hebei, Shandong, and Henan provinces. It can be divided further into three wheat-growing areas: the northern winter wheat area (Z1), the north Huanghuai winter wheat area (Z2) and the south Huanghuai winter wheat area (Z3).

The goal of wheat breeding is related to the development, orientation, and practical demand of the market economy (Han et al., 2017). As the breeding objectives change at the different stages, the traits of wheat varieties bred and marketed at the stages are different. Analyzing and summarizing the trend patterns and directions of wheat varieties can provide guidance for subsequent wheat breeding and production. At present, there have been a number of reports on the evolution of agronomic traits of wheat varieties in different wheat areas. For instance, in a report on winter wheat in America great plains, the authors investigated the yield attributes of 14 varieties introduced between 1873 and 1994 (Donmez et al., 2001). Their results suggested that the reduction of height, early maturity, and resistance to lodging contributed to the upward trend of yield. An analysis on winter wheat cultivars planted in France from 1946 to 1992 showed that the decrease of height played the most important role in winter wheat yield, followed by the ability to use the total above-ground biomass to produce more grains (Brancourt-Hulmel et al., 2003). An experiment using 117 winter wheat cultivars released between 1920 and 2000 to appraise the changes in various agronomic traits in Chile found that with the change of certification time, the harvest index increased and the plant height decreased gradually (Matus et al., 2012). The study on the evolution of wheat cultivars released in China also indicated increasing grain weight and yield with decreasing plant height and maximum tillers (Song et al., 2013).

The number of wheat varieties used in the aforementioned research ranged from dozens to hundreds. And the research objects mostly involved yield and yield-related elements. In many countries, distinctness, uniformity, and stability are the prerequisites for a new plant variety to obtain protection and registration (Deng et al., 2019). To meet the DUS criterion, a new wheat variety must be shown clearly different from all other varieties of common knowledge, reasonably uniformity and stability based on several of phenotypic characteristics (Wang et al., 2002). The phenotypic characteristics are listed in wheat test guidelines and range from morphological, phonological to agronomic traits. In the above-mentioned studies, the authors used almost exclusively yield and yield-related traits. Studies based on the DUS testing characteristics are reported rarely. Summarizing and analyzing the characteristics related to DUS

testing can not only evaluate the ability of these characteristics to distinguish varieties, but also provide supplementary information for new variety breeding (Wang et al., 2002; Reddy et al., 2009; Singh et al., 2012; Jones et al., 2013; Pilarczyk et al., 2015). In this study, 195 wheat varieties of the Huang-Huai-Hai origin were used as materials and 32 characteristics listed in China national test guidelines for wheat were observed and analyzed. The artificial selection trend of the varieties was analyzed in chronological order. We hope the results from this study can play a guiding role in the selection and cultivation of wheat varieties from the aspect of new variety approval.

MATERIALS AND METHODS

Plant Materials

The 195 wheat varieties used in this study were released in the Huang-Huai-Hai region in the past 20 years (**Supplementary Table 1**). The information on release time, suitable planting areas, and pedigrees were derived from the Data Platform of China Seed Industry (<http://202.127.42.145/home/manageorg>). There were 34, 79, and 82 varieties from the northern winter wheat area, the north Huanghuai winter wheat area and the south Huanghuai winter wheat area, respectively.

Method of Examination for DUS Testing Characteristics

The materials were planted in the experimental base of the Crop Research Institute, Shandong Academy of Agricultural Sciences (Jinan, China) in October 2019 and 2020. The experimental design was a 6-row area with a row length of 6 m and a row spacing of 0.25 m, with two repetitions.

The characteristics investigated were those listed in the national DUS test guidelines for common wheat, including a total of 32 qualitative characteristics (QL), pseudo-qualitative characteristics (PQ), and quantitative characteristics (QN). The detailed information on the 32 characteristics were listed in **Table 1**. The observations were made in accordance with the wheat test guidelines. The observation methods of characteristics included visual assessment by a single observation of a group of plants or parts of the plant (VG), measurement of a number of individual plants or parts of plants (MS), and a single measurement of a group of plants or parts of plants (MG). The corresponding codes of visually assessed characteristics and expression status were recorded according to the guidelines. For each measured characteristic, such as flag leaf length, flag leaf width, plant height, ear length, spikelet number per ear, sterile spikelet number per ear, and grain number per ear, at least 20 typical strains were selected, measured, and recorded one by one for each variety. The grain volume weight was measured by HGT-1000A in accordance with the guidelines (GB/T 5498-2013).

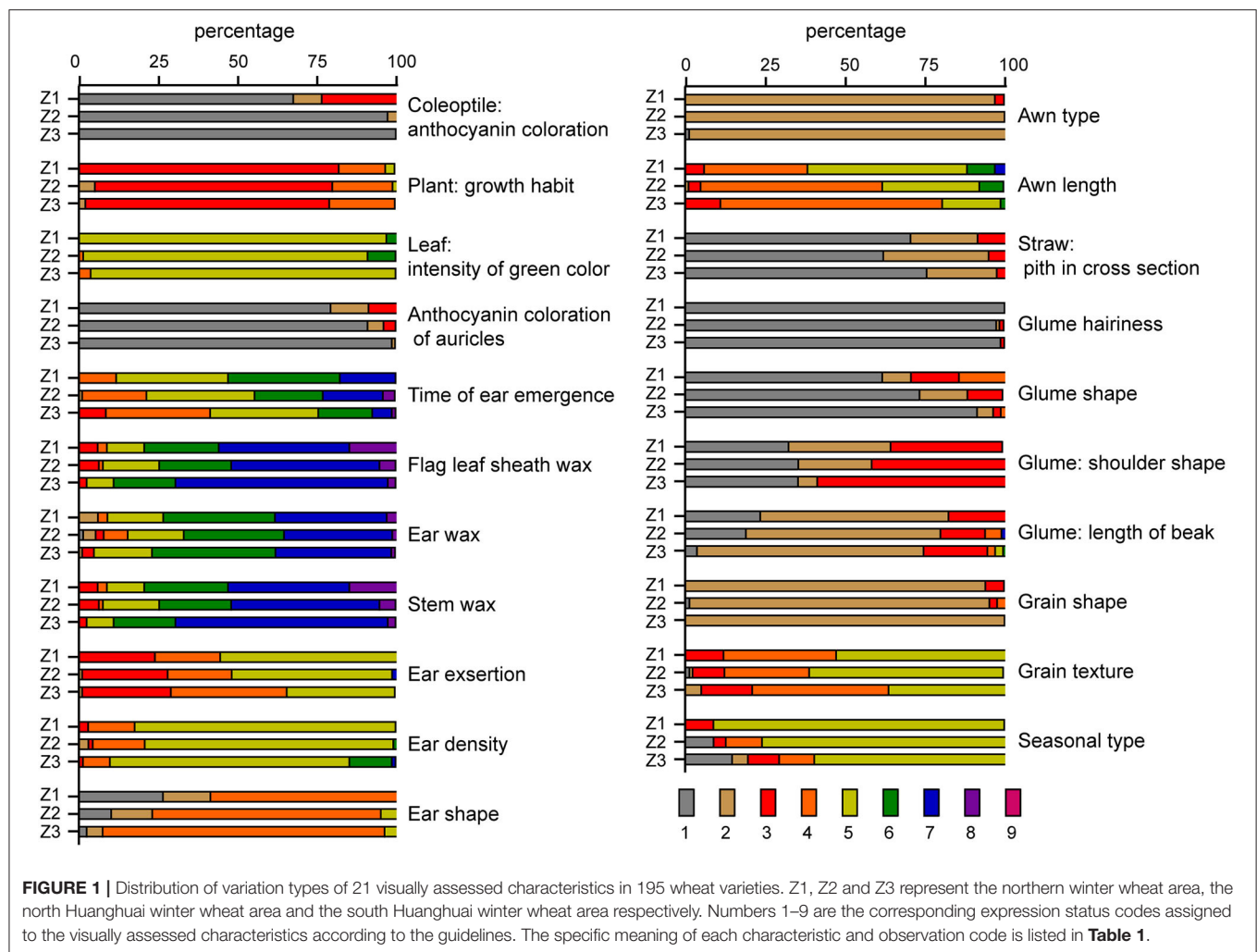
Statistical Analysis of Phenotypic Data

For the visually assessed characteristics, the observed codes across 2 years were sequentially corrected by the standard varieties. For each measured characteristic, means of 40 typical strains aggregated over 2 years were used for the statistical analysis. The data of each characteristic was analyzed by

TABLE 1 | The information on wheat DUS testing characteristics used in this study.

Characteristics	Character code	Type of expression	Method of observation	States and code of expression
Coleoptile: anthocyanin coloration (CAC)	char1	QN	VG	absent or very weak (1); weak (2); strong (3)
Plant: growth habit (PGH)	char2	QN	VG	Erect (1); semi erect (3); prostrate (5)
Leaf: intensity of green color (LIGC)	char3	QN	VG	light (3); medium (5); dark (7)
Anthocyanin coloration of auricles (ACA)	char5	QN	VG	absent or very weak (1); weak (2); strong (3)
Time of ear emergence (TEE)	char6	QN	VG	very early (1); early (3); medium (5); late (7); very late (9)
Flag leaf sheath wax (FLSW)	char7	QN	VG	absent or very weak (1); weak (3); medium (5); strong (7); very strong (9)
Ear wax (EW)	char8	QN	VG	absent or very weak (1); weak (3); medium (5); strong (7); very strong (9)
Stem wax (SW)	char9	QN	VG	absent or very weak (1); weak (3); medium (5); strong (7); very strong (9)
Flag leaf length (FLL)	char10	QN	MS	short (3); medium (5); long (7)
Flag leaf width (FLW)	char11	QN	MS	narrow (3); medium (5); broad (7)
Ear exertion (EE)	char12	PQ	VG	very short (1); short (3); medium (5); long (7); very long (9)
Plant height (PH)	char13	QN	MS	very low (1); low (3); medium (5); high (7); very high (9)
Ear length (EL)	char14	QN	MS	very short (1); short (3); medium (5); long (7); very long (9)
Spikelet number per ear (SNPE)	char15	QN	MS	few (3); medium (5); many (7)
Sterile spikelet number per ear (SSNPE)	char16	QN	MS	absent or very few (1); few (3); medium (5); many (7)
Grain number per ear (GNPE)	char17	QN	MS	very few (1); few (3); medium (5); many (7); very many (9)
Ear density (ED)	char18	QN	VG	lax (3); medium (5); dense (7)
Ear shape (ES)	char19	PQ	VG	fusiform (1); conical (2); ellipse (3); rectangle (4); stick shape (5); branched (6)
Awn type (AT)	char20	PQ	VG	absent (1); straight (2); curved (3)
Awn length (AL)	char21	QN	VG	very short (1); short (3); medium (5); long (7); very long (9)
Awn color (AC)	char22	QL	VG	light yellow (1); red (2); black (3)
Stem color (SC)	char23	QL	VG	yellow (1); purple (2)
Straw: pith in cross section (SPCS)	char24	QN	VG	absent or thin (1); medium (2); thick or filled (3)
Glume color (GLC)	char25	PQ	VG	yellow (1); red (2); black (3)
Glume hairiness (GLH)	char26	QN	VG	less (3); medium (5); multiple (7)
Glume shape (GLS)	char27	PQ	VG	ovate (1); near circular (2); ellipse (3); oblong (4)
Glume: shoulder shape (GLSS)	char28	QN	VG	inclined shoulder (1); square shoulder (2); shoulder (3)
Glume: length of beak (GLLB)	char29	QN	VG	very short (1); short (3); medium (5); long (7); very long (9)
Grain shape (GS)	char30	PQ	VG	oblong (1); oval (2); ellipse (3); near circular (4)
Grain texture (GT)	char33	QN	VG	silty (1); semi-horny (3); horny (5)
Seasonal type (ST)	char34	PQ	VG	spring type (1); partial spring type (2); semi-winter type (3); partial winter type (4); winter type (5)
Grain volume weight (GVW)	char41	QN	MG	very low (1); low (3); medium (5); high (7); very high (9)

QL, Qualitative characteristic; QN, Quantitative characteristic; PQ, Pseudo-qualitative characteristic; MS, measurement of a number of individual plants or parts of plants; MG, single measurement of a group of plants or parts of plants; VG, visual assessment by a single observation of a group of plants or parts of plants.



SPSS 21.0 (IBM Corp., Armonk, NY) for descriptive statistics and correlation. According to the release time, the wheat varieties were divided into four groups to analyze the artificial selection trends of DUS testing characteristics. The phenotypic distances of wheat varieties were carried out through R packages ‘dist’: Distance Matrix Computation. First, the data for each characteristic were standardized by dividing by the sample SD to remove the effect of dimensionality. Then, the various distances were calculated according to Euclidean and clustered using Ward’s clustering algorithm. The graphics were drawn by the ‘ggplot2’ package in R software or GraphPad Prism to visualize the results. The dendrogram was drawn and beautified using the online tool EvolView (<https://www.evolgenius.info/evolview/>).

RESULTS

Observation and Analysis of DUS Testing Characteristics

In total, 32 characteristics of 195 varieties from the three wheat-growing areas in the Huang-Huai-Hai region were observed and analyzed. There was only one state of expression observed

for characteristics awn color, stem color, and glume color in all the tested 195 varieties. Therefore, they would not be analyzed further. The results of the remaining 21 visually assessed characteristics were shown in **Figure 1**. For the straw pith in cross-section and glume shoulder shape, varieties from the three wheat-growing areas all had the same spectrum of expression states but different distribution frequencies for these states. For the other 19 visually assessed characteristics, varieties from the three wheat-growing areas showed clear differences in both the spectrum of expression states and distribution frequencies for these states. For the coleoptile anthocyanin coloration, leaf intensity of green color, anthocyanin coloration of auricles, awn type, and glume hairiness, varieties from the three wheat-growing areas shared the same major state of expression and had <3 states of expression. There were four or more states of expression for the time of ear emergence, flag leaf sheath wax, ear wax, stem wax, ear density, awn length, glume shape, and seasonal type in all three groups of varieties. Overall, the 195 varieties from the three wheat-growing areas displayed the rich genetic diversity.

The descriptive statistical results of the eight measured characteristics for varieties from three growing areas were

TABLE 2 | Statistical analysis results of eight measured characteristics of 195 wheat varieties.

Characteristics	Suitable planting areas	Mean ^a	STDEV	Min.	Max.	Range	C.V./%
Flag leaf length (FLL)	Z1	19.76a	2.6	14.5	28.13	13.64	13.14
	Z2	19.66a	2.37	14.64	28.02	13.38	12.04
	Z3	20.38a	2.25	13.78	25.89	12.11	11.06
Flag leaf width (FLW)	Z1	1.75b	0.2	1.42	2.08	0.66	11.19
	Z2	1.82b	0.2	1.2	2.51	1.31	11.16
	Z3	1.91a	0.23	1.54	3.01	1.47	11.98
Plant height (PH)	Z1	78.66a	7.53	65.67	98.33	32.67	9.57
	Z2	72.43b	7.43	53	97.67	44.67	10.26
	Z3	65.49c	6.09	50.67	79.35	28.68	9.3
Ear length (EL)	Z1	9.07a	0.77	7.38	11.83	4.46	8.45
	Z2	9.06a	0.92	7.04	12.27	6.35	10.15
	Z3	8.85a	0.65	7.51	11.9	4.39	7.29
Spikelet number per ear (SNPE)	Z1	18.02a	1.28	16	20.85	4.85	7.08
	Z2	18.18a	1.41	16.42	20.65	10.18	7.64
	Z3	18.37a	1.06	16.2	21.25	5.05	5.75
Sterile spikelet number per ear (SSNPE)	Z1	0.75b	0.47	0	1.75	1.75	63.3
	Z2	0.66b	0.51	0	2.1	2.1	76.93
	Z3	0.86a	0.45	0.05	1.95	1.9	52.05
Grain number per ear (GNPE)	Z1	50.39ab	4.71	43.8	64.55	20.75	9.35
	Z2	52.95a	6.14	38.8	73.3	34.5	11.6
	Z3	53.17a	4.49	40.05	64.8	24.75	8.44
Grain volume weight (GVW)	Z1	783.79a	39.56	709	838	129	5.05
	Z2	798.35a	31.57	706	842	136	3.95
	Z3	795.69a	22.35	705	833	128	2.81

^aThe various lowercase letters following the data represent significant differences at a 0.05 level based on one-way ANOVA.

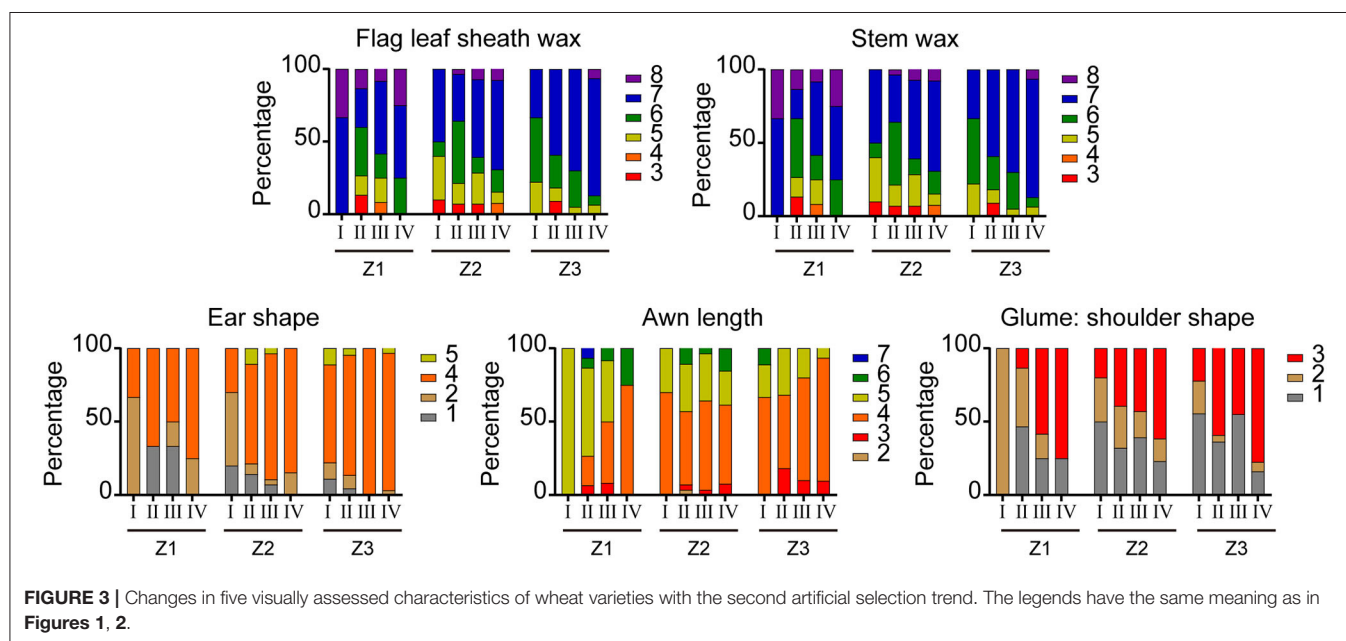
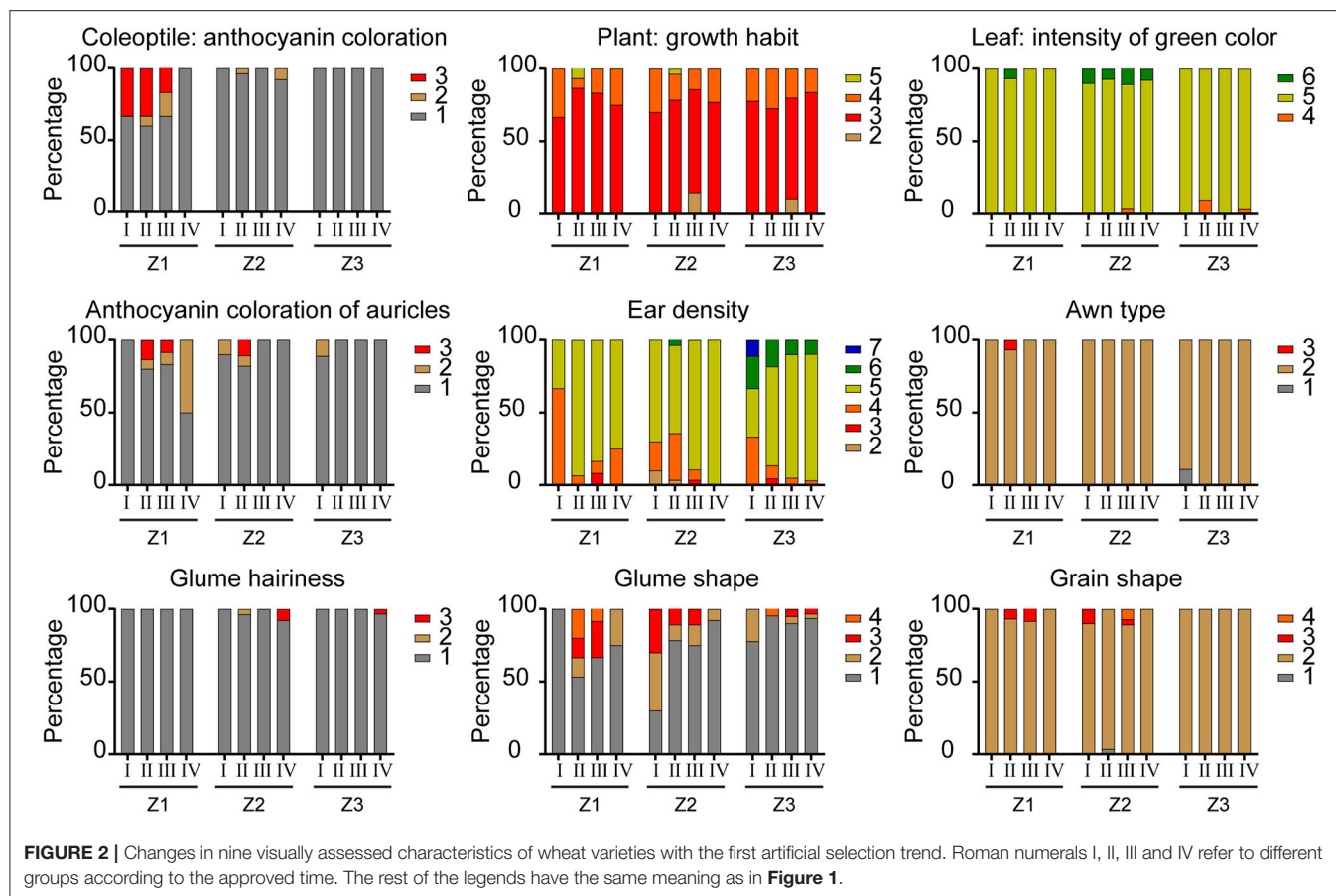
shown in **Table 2**. The mean, range, and coefficient of variation (CV) of the characteristic for each growing area differed from each other. Sterile spikelet number per ear had the highest CV among the eight characteristics with the CV of the varieties in Z2 being the largest. Grain volume weight had the smallest CV among the eight characteristics, with the varieties from Z3 showing the lowest CV among the three wheat-growing areas. The differences between the three growing areas on flag leaf width, plant height, sterile spikelet number per ear, and grain number per spike were significant ($p < 0.05$).

The Artificial Selection Trend of Tested Materials

Varieties in this study were divided into four groups according to their approved time. Groups I, II, III, and IV each consisted of 22 varieties registered before 2000, 65 varieties registered between 2001 and 2010, 60 varieties registered between 2011 and 2015, and 48 varieties registered since 2016. The information on registration time and suitable planting areas for the varieties is listed in **Supplementary Table 1**.

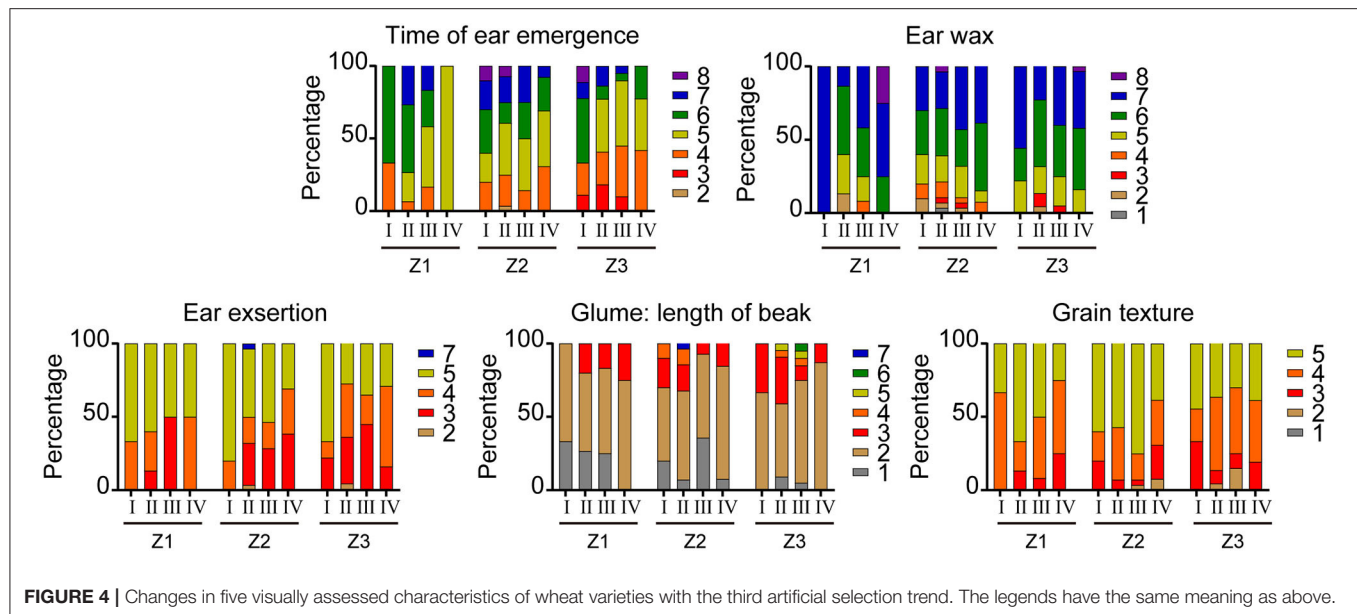
The artificial selection trends of DUS testing characteristics of varieties from the three growing areas released at different stages were analyzed. Among the 195 varieties, the straw pith

in cross-section and seasonal type had great differences between the three wheat-growing areas. With the change of years, the “absent or thin” type of straw pith became the dominant one and the spring type became the only type for varieties in the northern winter wheat area (Z1), while the expression states remained diversified for varieties in the north Huanghuai winter wheat area (Z2) and the south Huanghuai winter wheat area (Z3) on the two characteristics. The other 19 visually assessed characteristics in varieties of the three wheat areas showed similar artificial selection trends, which could be classified into three types. For type one (**Figure 2**), one state of expression dominated in all stages of registration. This type included nine characteristics: coleoptile anthocyanin color, plant growth habit, leaf intensity of green color, anthocyanin coloration of auricles, ear density, awn type, glume hairiness, glume shape, and the grain shape. For type two (**Figure 3**), the proportion of a certain state of expression increased gradually over the years and changed in a specific direction. Such characteristics included flag leaf sheath wax, stem wax, ear shape, awn length, and glume shoulder shape. The artificial selection trends of the first and second types were direct or indirect results of purposeful selection by breeders in the breeding process. For type three (**Figure 4**), a number of expression states (three or more) existed for the characteristics and the proportions of the expression states did not change significantly over



the years. The characteristics of type three included time of ear emergence, ear wax, ear exertion, glume length of beak, and grain texture. The stable distribution pattern of

states of expression over years of type three characteristics suggested that they were not the main factors in the breeding selection process.



The statistical results and trend diagrams for the eight measured characteristics of registered varieties across different periods and wheat regions were shown in **Figure 5**. For flag leaf length and sterile spikelet number per ear, varieties in both Z1 and Z2 showed a downward trend, although the trend for varieties in Z1 was more obvious. Flag leaf width and grain number per spike showed a growing trend, however, the increase was more pronounced for varieties in Z1 than varieties in Z2 and Z3. Ear length remained more or less stable in the three wheat-growing areas. Spikelet number per ear showed a similar downward tendency in Z1 and Z3, but the decreasing trend in Z1 was more significant. In all three areas, plant height had been declining while grain volume weight had been increasing. However, the strength of the trends over the years for varieties from the three areas were significantly different. The plant height for varieties in Z1 showed the fastest rate of decrease, while the grain volume weight increased more rapidly for varieties in Z3. For the varieties registered after 2016, the mean plant height for varieties in Z3 were significantly lower than both varieties in Z1 and Z2, while the mean grain volume weight for varieties in Z3 was the highest among varieties from the three areas.

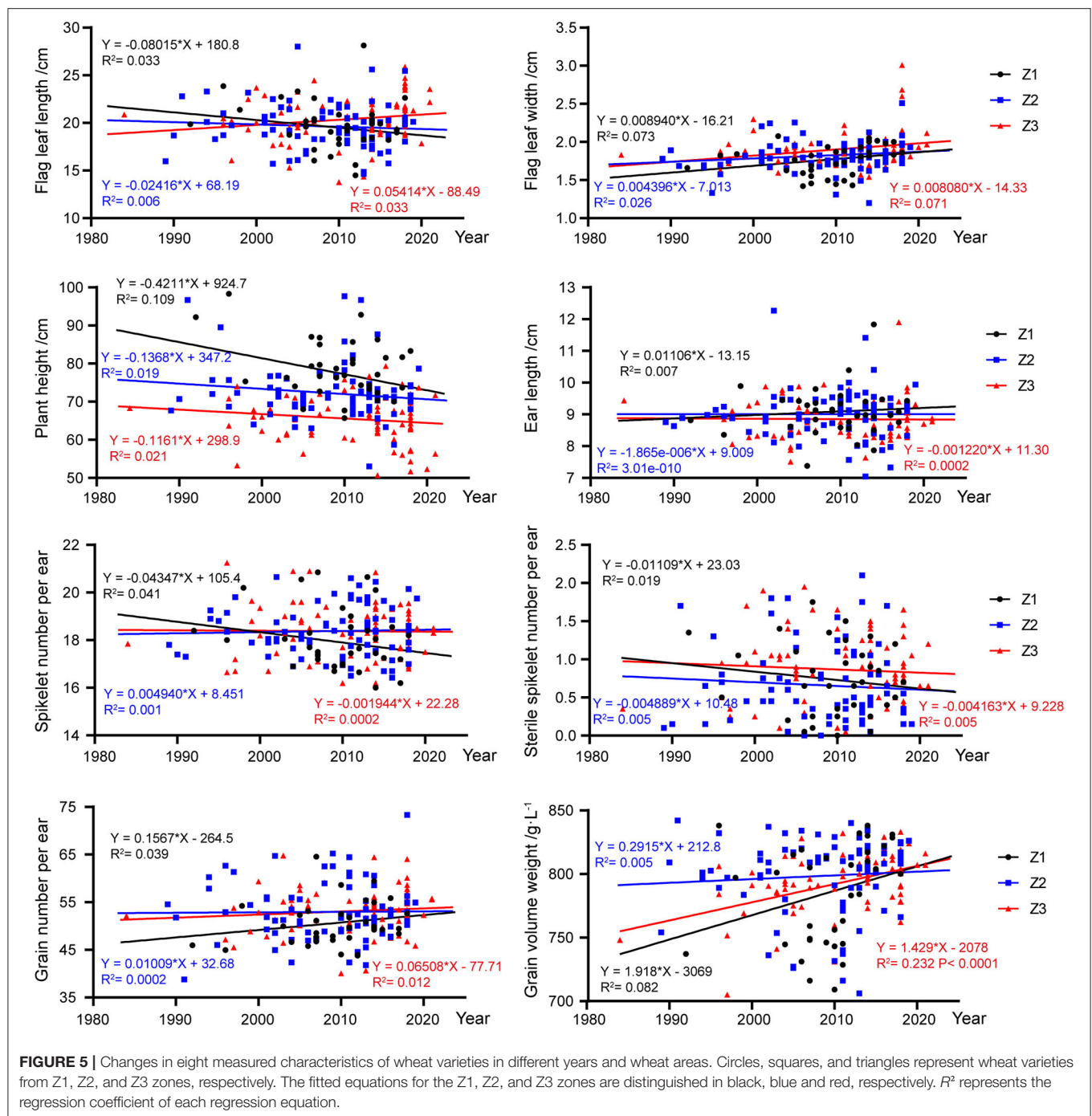
Correlations Between DUS Testing Characteristics

Correlation analysis among DUS testing characteristics of the 195 varieties was conducted and the correlation results were listed in **Supplementary Table 2**. The correlation coefficients among the three wax-related characteristics of flag leaf sheath wax, ear wax, and stem wax reached significant level. Flag leaf length had a significant positive correlation with flag leaf width. For the seven ear-related characteristics, ear density was positively correlated with ear shape and the number of

grains per ear. Ear length was negatively correlated with ear shape but positively correlated with spikelet number per ear. Grain number per ear was positively correlated with ear length, spikelet number per ear, ear density, and ear shape, and was negatively correlated with sterile spikelet number per ear. For the remaining characteristics, the correlation coefficients among awn-related characteristics (awn type and awn length), glume-related characteristics (glume shape and glume length of beak) and grain-related characteristics (grain shape, grain texture, and grain volume weight) did not reach significant levels. In the combination of characteristics with significant relationships, the correlation between flag leaf sheath wax and stem wax was the strongest with a correlation coefficient of 0.99. Most of the remaining correlation coefficients ranged from 0 to 0.4, indicating that there was only a weak relationship between the relevant characteristics.

Cluster Analysis Based on DUS Testing Characteristics and Pedigrees

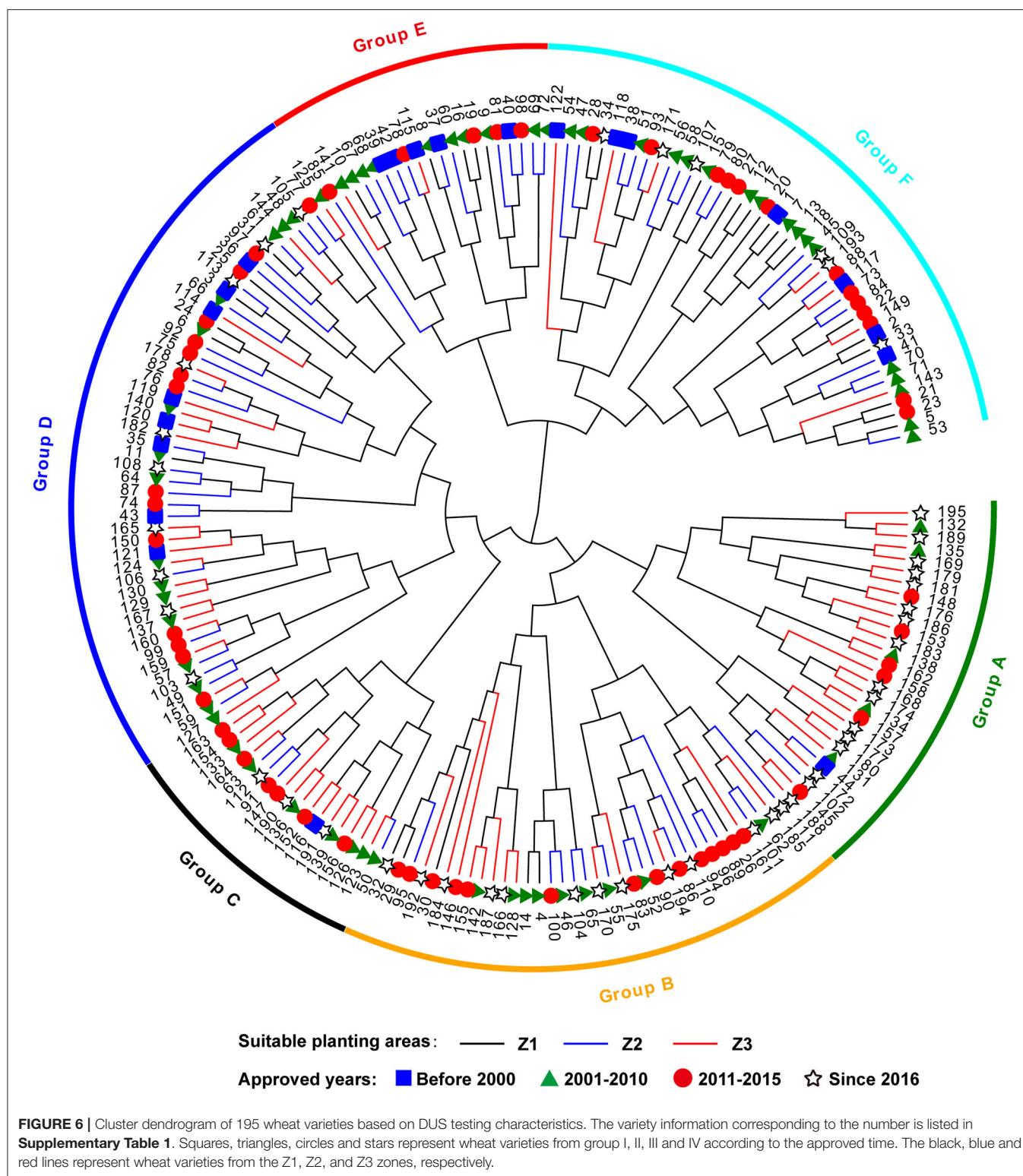
The 195 wheat varieties were divided into six groups according to the clustering dendrogram constructed from the DUS testing characteristics (**Figure 6**). Group A included 28 wheat varieties that were mainly suitable for planting in the south Huanghuai winter wheat area (Z3) and approved since 2016. Group B contained 37 wheat varieties mainly from the north Huanghuai winter wheat area (Z2) and the south Huanghuai winter wheat area (Z3). And most of the wheat varieties in Group B were released between 2011 and 2015. Group C was the smallest one, including 17 wheat varieties mainly from Z3, and released between 2000 and 2015. Group D was the largest one with 50 wheat varieties mainly from Z2 and Z3. Group E had 20 wheat varieties from the north Huanghuai winter wheat area and the approval was before 2010. Group F contained the



remaining 43 wheat varieties from the northern winter wheat area (Z1) and the north Huanghuai winter wheat area (Z2). The approval time of wheat varieties in Groups D and F was mainly before 2015. From the clustering results in **Figure 6**, the wheat varieties in Z1 were the farthest from those in Z3. And most of the wheat varieties on the same clade were also approved during the same period. The results of this study showed that artificial selection of characteristics was consistent

with ecological adaptation and the breeding process to a certain extent.

According to the information of pedigrees obtained from the Data Platform of China Seed Industry **Supplementary Table 1**), we found that the parents of most wheat varieties in Groups A and C were Zhoumai and Yumai series materials from the Henan Province, among which Zhoumai16 and Zhoumai22 were the most widely used parent materials. The wheat varieties in



Group B were mostly derived from parent lines from Shandong and Hebei provinces. Group D contains the largest number of materials and the most extensive sources. Their parental materials came from Hebei, Shandong, and Henan provinces.

Nonetheless, the materials with similar sources are more closely branched in the clustering results. Therefore, the clustering results of DUS testing characteristics were also consistent with pedigree analysis.

DISCUSSION

At present, most research reports on the evolution of agronomic traits of wheat varieties focused mainly on the yield and yield-related components or quality traits. In contrast to those, the characteristics specified in the DUS test guidelines of wheat were used in this study for observation and data analysis. The wheat materials used in this study were those bred and released in the Huang-Huai-Hai region in China in the last 20 years. Most characteristics of these wheat varieties were rich in variation. The results indicated that the wheat materials have high-genetic diversity in these phenotypic characteristics. As a result, the wheat materials can be effectively distinguished and evaluated according to these DUS testing characteristics. However, there was a little difference in awn color, stem color, and glume color among the 195 wheat varieties. It could be explained that these characteristics were irrelevant to the current-breeding objectives.

According to the analysis, results of 195 varieties based on the sources of different wheat-growing areas, it was found that straw pith and seasonal type had large differences between the northern winter wheat zone (Z1), the north Huanghuai winter wheat zone (Z2), and the south Huanghuai winter wheat zone (Z3). The artificial selection trends of the other 19 visually assessed characteristics were similar in the three ecological zones, which could be grouped into three types except for straw pith in cross-section and seasonal type. Among the eight measured characteristics, wheat varieties in the three zones showed the same trends in flag leaf width, plant height, spikelet number per ear, and grain volume weight. There was a declining trend for plant height and an increasing trend for grain volume weight. It has been previously reported that linear regression is suitable for quantitative characteristics to predict changing trends (Jones et al., 2013). Regrettably, the *P*-values between the measured characteristics and the fitted linear equations did not reach a very significant level in this study, except for the trend of grain volume weight for varieties in the south Huanghuai winter wheat zone. We speculate that this may be related to the number of varieties in this study, so other approved wheat varieties remain to be tracked.

Previous studies have shown that abundant waxes on the plant in wheat can improve the resistance to drought (Elham et al., 2012; Dai et al., 2016; Liu et al., 2017) and other biotic and abiotic stresses (Koch et al., 2006; Nawrath, 2006; Yeats and Rose, 2013). The waxes can also improve photosynthesis, reduce transpiration, improve water use efficiency, and ultimately improve grain yield and biological yield (Wang et al., 2015; Xu et al., 2016; Yang et al., 2018; Liu et al., 2020a). Therefore, the wax intensity on flag leaf, stem, ear, and other organs can be used as an indicator for drought-resistant and stress-resistant wheat cultivars. The results showed that the wax on flag leaf sheath and stem showed a gradually increasing trend over time (Figure 3), which suited the needs for drought and stress resistance of wheat. Among the components of wheat yield, fertile ears per hectare have the highest correlation with the yield, followed by 1,000-grain weight (Shi et al., 2009). Studies have shown that traits such as ear length, number of spikelets per ear, number of grains per ear, and volume weight are positively correlated with wheat

yield (Zhou et al., 2007; Sadras and Rebetzke, 2013). Reducing the number of sterile spikelets per ear can increase the number of grains per spike, and thus, increase wheat yield. Similarly, it has been reported that plant height has a significant correlation with wheat yield, and reducing plant height can improve harvest index and enhance lodging resistance (Rebetzke and Richards, 2000; Rebetzke et al., 2011). But an unrestricted reduction in plant height can also limit the improvement in biological yield, so both plant height overabundance and overweigh can cause a reduction in yield. The characteristics we used in this study were all derived from the DUS test guidelines and the yield was not measured directly. According to the changing trend of tested characteristics, the length and width of flag leaf, the number of grains per ear, and the grain volume weight of the certified varieties showed an overall upward trend, while the plant height decreased gradually. Our results were consistent with those from studies of cultivars conducted in the previous studies (Donmez et al., 2001; Brancourt-Hulmel et al., 2003; Matus et al., 2012; Song et al., 2013; Liu et al., 2020b). Therefore, it was speculated that wheat yield would be increased by adjusting the plant height, reducing the number of sterile spikelets per ear, and increasing the number of grains per ear, which was in accordance with the mainstream goal of current breeding.

CONCLUSION

In this study, 195 wheat varieties in the Huang-Huai-Hai region were used to study their artificial selection trend evaluated by thirty-two DUS testing characteristics. From the perspective of testing for new wheat variety, awn color, stem color, and glume color, is contributive in enriching the genetic diversity of varieties. The straw pith in cross-section and seasonal type had great differences between the three wheat-growing areas under the artificial selection. For the measured characteristics, flag leaf width, plant height, and sterile spikelet number per ear were more closely related to ecological fitness than other measured characteristics. The cluster analysis indicated that the DUS testing characteristics can effectively cluster materials from the same or similar areas, materials from the same approved period, and materials with closer kinship in the same clade. Our results in this study could provide guidance for screening similar varieties and determining the distinctness of applied varieties in DUS testing.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

HZ and RL designed the research and corrected the manuscript. LW performed the main analyses, prepared figures, and wrote the manuscript. YZ helped in data analysis and revised the manuscript. LD and HL conducted experiments and collected

data. MW and HW performed the experiments and supervised the study. All the authors read and approved the final version of the manuscript.

FUNDING

This work was supported by grants from the Study on Molecular Marker Assisted Evaluation of Distinctiveness, Uniformity, and Stability of New Wheat Varieties (2015YQN14), the Funds for

DUS Testing and Maintenance of Known Variety Library in 2021, and the efficient and accurate detection of crop varieties and seeds (CXGC2022C01).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.898102/full#supplementary-material>

REFERENCES

- Brancourt-Hulmel, M., Doussinault, G., Lecomte, C., Bérard, P., Buanec, B. L., and Trotter, M. (2003). Genetic improvement of agronomic traits of winter wheat cultivars released in France from 1946 to 1992. *Crop Sci.* 43, 37–45. doi: 10.2135/cropsci2003.0037
- Dai, S., Guo, J., Xu, W., Zhao, S. J., and Song, J. M. (2016). Biosynthesis and regulation of cuticular wax and its effects on drought resistance of wheat. *Plant Physiol. J.* 52, 979–988. doi: 10.13592/j.cnki.ppj.2016.0135
- Deng, C., Han, R. X., Yang, X. H., and Tang, H. (2019). Stability of agricultural plant varieties in China. *China Seed Ind.* 286, 10–11.
- Donmez, E., Sears, R. G., Shroyer, J. P., and Paulsen, G. M. (2001). Genetic gain in yield attributes of winter wheat in the great plains. *Crop Sci.* 41, 1412–1419. doi: 10.2135/cropsci2001.4151412x
- Elham, F., Khavari-Nejad, R. A., Salekdeh, G. H., and Najafi, F. (2012). Evaluation of cuticular wax deposition, stomata and carbohydrate of wheat leaves for screening drought tolerance. *Adv. Environ. Biol.* 6, 4035–4040.
- Han, J. P., Yu, X. F., Chang, J. L., Yang, G. X., and He, G. Y. (2017). Overview of the wheat genetic transformation and breeding status in China. *Methods Mol. Biol.* 1679, 37–60. doi: 10.1007/978-1-4939-7337-8_3
- Jones, H., Norris, C., Smith, D., Cockram, J., Lee, D., O'Sullivan, D. M., et al. (2013). Evaluation of the use of high-density SNP genotyping to implement UPOV Model 2 for DUS testing in barley. *Theor. Appl. Genet.* 126, 901–911. doi: 10.1007/s00122-012-2024-2
- Koch, K., Barthlott, W., Koch, S., Hommes, A., Wandelt, K., Mamdouh, W., et al. (2006). Structural analysis of wheat wax (*Triticum aestivum*, c.v. "Naturastar" L.): from the molecular level to three dimensional crystals. *Planta* 223, 258–270. doi: 10.1007/s00425-005-0081-3
- Liu, X. W., Feakins, S. J., Dong, X. J., Xue, Q. W., Marek, T., Leskovar, D., et al. (2017). Experimental study of leaf wax n-alkane response in winter wheat cultivars to drought conditions. *Org. Geochem.* 113, 210–223. doi: 10.1016/j.orggeochem.2017.07.020
- Liu, X. Y., Li, X. H., Zheng, X. W., Qiao, L., Zhao, J. J., Ge, C., et al. (2020a). Evolution and relevant analysis of agronomic characters of wheat in Shanxi province. *J. Agric. Sci. Technol.* 22, 14–23.
- Liu, X. Y., Wu, H. Q., Shi, S. D., Li, C. L., Wang, Y., and Wang, Z. H. (2020b). Cloning and functional analysis of epidermal wax fatty alcohol synthesis gene *TaFAR10* in common wheat. *J. Triticeae Crops* 40, 21–29.
- Matus, I., Mellado, M., Pinares, M., Madariaga, R., and Pozo, A. D. (2012). Genetic progress in winter wheat cultivars released in Chile from 1920 to 2000. *Chil. J. Agric. Res.* 72, 303–308. doi: 10.4067/s0718-58392012000300001
- Nawrath, C. (2006). Unraveling the complex network of cuticular structure and function. *Curr. Opin. Plant Biol.* 9, 281–287. doi: 10.1016/j.pbi.2006.03.001
- Pilarczyk, W., Kowalczyk, B., and Bakińska, E. (2015). Comparison of uniformity decisions in DUS testing for full and reduced numbers of measurements. *Biom. Lett.* 52, 47–53. doi: 10.1515/bile-2015-0005
- Rebetzke, G. J., Ellis, M. H., Bonnett, D. G., Condon, A. G., Falk, D., and Richards, R. A. (2011). The *Rht13* dwarfing gene reduces peduncle length and plant height to increase grain number and yield of wheat. *Field Crops Res.* 124, 323–331. doi: 10.1016/j.fcr.2011.06.022
- Rebetzke, G. J., and Richards, R. A. (2000). Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. *Crop Pasture Sci.* 51, 235–246. doi: 10.1071/AR99043
- Reddy, D., Audilakshmi, S., and Seetharama, N. (2009). Genetic variability and divergence for DUS testing traits in sorghum (*Sorghum bicolor*). *Indian J. Agric. Sci.* 79, 286–290.
- Sadras, V. O., and Rebetzke, G. J. (2013). Plasticity of wheat grain yield is associated with plasticity of ear number. *Crop Pasture Sci.* 64, 234–243. doi: 10.1071/CP13117
- Shi, M. F., An, L. L., Xing, C. P., Han, D. C., and Ding, G. Q. (2009). Effect of ear style to yield composition in winter wheat. *J. Shanxi Agric. Sci.* 36, 38–40.
- Singh, B., Chaubey, T., Upadhyay, D. K., and Jha, A. P. (2012). Morphological characterization for DUS testing of cabbage (*Brassica oleracea* var. *capitata* L.) cultivars. *Progressive Hortic.* 44, 32–36.
- Song, J. M., Dai, S., Li, H. S., Cheng, D. G., Liu, A. F., Cao, X. Y., et al. (2013). Evolution of agronomic and quality traits of wheat cultivars released in Shandong province recently. *Sci. Agric. Sin.* 46, 1114–1126.
- Wang, Y., Wang, M. L., Sun, Y. L., Hegebarth, D., Li, T. T., Jetter, R., et al. (2015). Molecular characterization of *TaFAR1* involved in primary alcohol biosynthesis of cuticular wax in hexaploid wheat. *Plant Cell Physiol.* 56, 1944–1961. doi: 10.1093/pcp/pcv112
- Wang, Y. R., Cui, Y. H., Nan, Z. B., and Wang, W. (2002). Characteristic selection and example variety determination of DUS testing guidelines of plant new variety. *Pratacultural Sci.* 2, 44–47. doi: 10.1088/1009-1963/11/5/313
- Xu, W., Shen, H., Guo, J., Yu, X. C., Li, X., Yang, Y. H., et al. (2016). Drought resistance of wheat NILs with different cuticular wax contents in flag leaf. *Acta Agron. Sin.* 42, 1700–1707. doi: 10.3724/SP.J.1006.2016.01700
- Yang, Y. H., Xiao, M. A., Zhang, Z. S., Guo, J., and Zhao, S. (2018). Effects of drought stress on photosynthetic characteristics of wheat near-isogenic lines with different wax contents. *Sci. Agric. Sin.* 51, 4241–4251. doi: 10.3864/j.issn.0578-1752.2018.22.003
- Yeats, T. H., and Rose, J. (2013). The formation and function of plant cuticles. *Plant Physiol.* 163, 5–20. doi: 10.1104/pp.113.222737
- Zhou, Y., He, Z. H., Chen, X. M., Wang, D. S., Zhang, Y., and Zhang, G. S. (2007). Genetic gain of wheat breeding for yield in northern winter wheat zone over 30 years. *Acta Agron. Sin.* 33, 1530–1535.

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Zheng, Duan, Wang, Wang, Li, Li 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) and the copyright owner(s) 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.



Genomic Analysis Provides Insights Into the Plant Architecture Variations in *in situ* Conserved Chinese Wild Rice (*Oryza rufipogon* Griff.)

Ziyi Yang^{1,2†}, Yilin Zhang^{3,5†}, Meng Xing^{1,2†}, Xiaowen Wang³, Zhijian Xu⁴, Jingfen Huang¹, Yanyan Wang^{1,2}, Fei Li¹, Yamin Nie¹, Jinyue Ge¹, Danjing Lou¹, Ziran Liu¹, Zhenyun Han¹, Yuntao Liang⁴, Xiaoming Zheng^{1,2}, Qingwen Yang^{1,2*}, Hang He^{3,5*} and Weihua Qiao^{1,2*}

¹ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ² National Nanfan Research Institute, Chinese Academy of Agricultural Sciences, Sanya, China, ³ State Key Laboratory of Protein and Plant Gene Research, School of Advanced Agriculture Sciences and School of Life Sciences, Peking University, Beijing, China, ⁴ Rice Research Institute, Guangxi Academy of Agricultural Sciences, Nanning, China, ⁵ Institute of Advanced Agricultural Sciences, Peking University, Weifang, China

OPEN ACCESS

Edited by:

Chuanzhi Zhao,
Shandong Academy of Agricultural
Sciences, China

Reviewed by:

Jian Sun,
Shenyang Agricultural University,
China
Muthusamy Ramakrishnan,
Nanjing Forestry University, China

*Correspondence:

Qingwen Yang
yangqingwen@caas.cn
Hang He
hehang@pku.edu.cn
Weihua Qiao
qiaoweihua@caas.cn

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 15 April 2022

Accepted: 24 May 2022

Published: 27 June 2022

Citation:

Yang Z, Zhang Y, Xing M, Wang X,
Xu Z, Huang J, Wang Y, Li F, Nie Y,
Ge J, Lou D, Liu Z, Han Z, Liang Y,
Zheng X, Yang Q, He H and Qiao W
(2022) Genomic Analysis Provides
Insights Into the Plant Architecture
Variations in *in situ* Conserved
Chinese Wild Rice (*Oryza rufipogon*
Griff.). *Front. Plant Sci.* 13:921349.
doi: 10.3389/fpls.2022.921349

In situ conserved wild rice (*Oryza rufipogon* Griff.) is a promising source of alleles for improving rice production worldwide. In this study, we conducted a genomic analysis of an *in situ* conserved wild rice population (Guiping wild rice) growing at the center of wild rice genetic diversity in South China. Differences in the plant architecture in this population were investigated. An analysis using molecular markers revealed the substantial genetic diversity in this population, which was divided into subgroups according to the plant architecture. After resequencing representative individuals, the Guiping wild rice population was compared with other *O. rufipogon* and *Oryza sativa* populations. The results indicated that this *in situ* conserved wild rice population has a unique genetic structure, with genes that were introgressed from aromatic and *O. sativa* ssp. *indica* and *japonica* populations. The QTLs associated with plant architecture in this population were detected via a pair-wise comparison analysis of the sequencing data for multiple DNA pools. The results suggested that a heading date-related gene (*DHD1*) might be associated with variations in plant architecture and may have originated in cultivated rice. Our findings provide researchers with useful insights for future genomic analyses of *in situ* conserved wild rice populations.

Keywords: Chinese wild rice, *in situ* conservation, genomic analysis, plant architecture variation, *OsDHD1*

INTRODUCTION

Asian cultivated rice (*Oryza sativa*) is one of the oldest staple food crops widely consumed by half of the global population, which is expected to grow by 25% and reach 10 billion over the next 30 years. Developing new rice varieties that are higher yielding and more nutritious than current cultivated varieties as well as resistant to pests and diseases and ideal for climate-smart agricultural production is critical for avoiding widespread malnutrition or starvation (Lee et al., 2019). Although the rice domestication history remains controversial, rice is thought to have been domesticated from the wild relative *Oryza rufipogon* in China approximately 9,000 years ago (Oka, 1988; Fuller et al., 2010; Civan et al., 2015; Wang et al., 2017). *Oryza rufipogon* grows in pantropical regions, including the diverse natural habitats in South China (Gao et al., 1996; Gao, 2004). Additionally, wild rice can be crossed directly with cultivated rice because they have the same AA genome, making *O. rufipogon*

a suitable genetic resource for enhancing rice characteristics. A key research milestone was the identification of the cytoplasmic male sterility genes in *O. rufipogon*. This enabled the development of three-line hybrid rice, which is a notable achievement in China, with worldwide implications.

China has one of the most abundant *O. rufipogon* resources worldwide. Specifically, *O. rufipogon* plants can be found in seven provinces and/or autonomous regions in China (i.e., Guangdong, Guangxi, Hainan, Yunnan, Hunan, Jiangxi, and Fujian), but their continued growth is at risk because of recent human activities and natural degradation (Xu et al., 2020). The *in situ* conservation of wild relatives of crops has been one of the key strategies for conserving biological diversity in China and elsewhere. Globally, China is the first country to implement the *in situ* conservation of wild rice. In the past two decades, there has been substantial progress in the *in situ* conservation of wild rice throughout China. Most of the endangered wild rice populations have been protected by both *in situ* and *ex situ* conservation programs (Xu et al., 2020). By the end of 2020, more than 30 national *in situ* conservation sites were constructed for wild rice in China (Yang et al., 2022). There is an urgent need for research that attempts to address many basic issues, including how to better develop and evaluate *in situ* conservation methods. Various molecular markers, such as simple sequence repeats (SSR) and nuclear intron targeting (InDel), were used to evaluate the genetic diversity in different crop wild relative species (Ramakrishnan et al., 2016). Research on the genetic diversity and population variation within a species may help clarify evolutionary processes and mechanisms, while also generating relevant information for designing appropriate and efficient *in situ* conservation strategies. However, analyses of genes and their introgression in or between *in situ* conserved populations remain limited.

The change from the creeping growth of cultivated rice ancestors to the erect growth of modern rice varieties is an example of the effect of domestication on plant architecture (Jin et al., 2008). Many plant architecture-related genes, such as *PROG1* and *LAZY1*, reportedly originated in *O. rufipogon* and have undergone natural variations in cultivated rice (Li et al., 2007; Tan et al., 2008). Most *O. rufipogon* plants exhibit creeping growth, but there are also some tilted and semi-erect types in some populations. In contrast, erect plants are rare in *O. rufipogon* populations. It is unclear whether erect wild rice plants are the result of the introgression of a gene from cultivated rice or a natural variation during the adaptation to environmental conditions.

In this study, we performed an extensive field and genomic investigation of a typical natural *O. rufipogon* population and its habitats within the *in situ* conservation sites at the center of wild rice genetic diversity in China. The plants in the selected population varied considerably in terms of plant architecture. The objectives of this study were as follows: (1) assess the genetic diversity within an *in situ* conserved population; (2) detect genes that were introgressed from other wild rice populations or local landraces; and (3) identify natural variations associated with plant architecture in the *in situ* conserved population. Our results provide useful information regarding plant architectural diversity

and new insights for optimizing the *in situ* conservation of wild rice in China.

MATERIALS AND METHODS

Population Sampling

To ensure the collected samples fully represented the genetic diversity of the population and to avoid collecting clones of the same individual, irregular grid lines were applied. Specifically, using an individual collection distance of about 5 m, a total of 184 randomly selected leaves were collected from their original habitat, Guiping *in situ* conservation site, placed in a plastic bag, and stored at -20°C . Collection of 30 samples in the mixing pool were performed also using a 5 m-distance rule. For the 20 re-sequenced samples, including 3 erect types samples, 4 semi-erect samples, 5 tilt samples, and 8 creep samples, same type individual collection distance of more than 20 m.

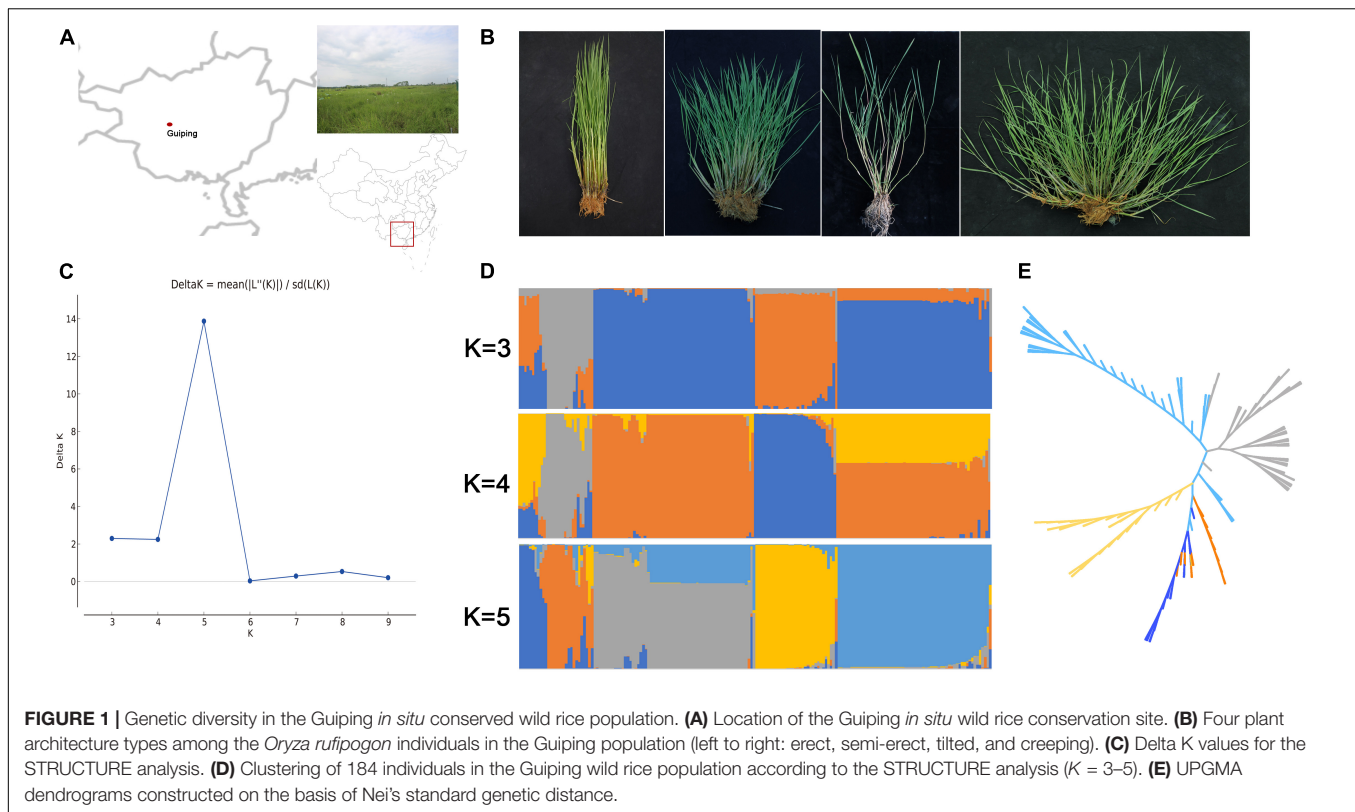
DNA Isolation and Polymerase Chain Reaction

The collected leaves were ground to a powder in liquid nitrogen using the SPEX Sample Prep Geno/Grinder. Genomic DNA was extracted according to the CTAB method (Edwards et al., 1991). A PCR amplification was completed in a 15- μL reaction volume that comprised 4.3 μL ddH₂O, 0.6 μL forward and reverse primers, 2 μL DNA, and 7.5 μL 2 \times Es Taq MasterMix. The PCR program was as follows: 95°C for 5 min; 35 cycles of 94°C for 30 s, 57°C for 30 s, and 72°C for 30 s; and 72°C for 2 min. The amplified products were separated in 6% denaturing polyacrylamide gels.

Statistical Analysis

The simple sequence repeat (SSR) data format was converted to readable input files for STRUCTURE, PowerMarker, and POPGENE. We used POPGENE 1.31 (Yeh et al., 1998) to calculate the following genetic diversity-related parameters: Shannon–Weaver information index (*I*), observed heterozygosity (*He*), and expected heterozygosity (*Ho*) (Nei, 1978). STRUCTURE was used to infer genetic clustering (K) according to a model-based clustering method (Falush and Pritchard, 2003). We evaluated K values from 2 to 9 by performing 10 independent runs for each K value; the model was repeated with a 10,000 burn-in period and 100,000 Monte Carlo Markov chain repetitions. CLUMPP (version 1.1) (Jakobsson and Rosenberg, 2007) was used to obtain the best clustering for each K value. The unweighted pair group method with arithmetic mean (UPGMA) (Liu and Muse, 2005) of PowerMarker was used to evaluate the relationships between populations according to Nei's standard genetic distance (Nei, 1972).

Sequences were aligned using the MUSCLE program in MEGA 7. A phylogenetic analysis involving nucleotide sequences was conducted using the maximum likelihood method, with 1,000 bootstrap replicates. A model-based estimation of the ancestry coefficient was performed using ADMIXTURE (Alexander et al., 2009). High-quality variations, including single



nucleotide polymorphisms (SNPs) and insertions/deletions (InDels), were identified by analyzing VCF files according to the following four criteria: (1) maximum missing rate < 0.2; (2) minimum read depth of 5; (3) biallelic position; and (4) genotype quality > 20. A haplotype network was constructed using the minimum spanning method implemented in POPART.

Pair-Wise Comparison Analysis for Multiple Pool-Seq (PCAMP)

We developed four DNA bulks from erect, semi-erect, tilted, and creeping Guiping plants at the *in situ* conservation site. The sample detection, library construction, library quality detection, and onboard sequencing were performed according to the standard Illumina protocols. After confirming the quality of the genomic DNA was sufficient for the subsequent analysis, the DNA was fragmented *via* a mechanical method. The ends of the fragmented DNA were repaired and a poly-A tail was added to the 3' end. After adding a sequencing adapter, a PCR amplification was performed and the resulting products were purified to complete the construction of the sequencing library. The high-quality library was sequenced using an Illumina system. We used the BWA software to align the short reads of the DNA bulks to the Nipponbare reference genome (MSU_v7.0) (Li and Durbin, 2009). The GATK software toolkit was used to detect SNPs (Yang et al., 2019). On the basis of the localization of the clean reads in the reference genome, Picard was used to eliminate redundant reads. The SnpEff software (Cingolani et al., 2012) was used to annotate mutations (SNPs and small InDels) and

predict the effects of the mutations. We calculated the Euclidean distance (ED) to identify the candidate regions associated with plant architecture. The ED was calculated using the following formula:

$$ED = \sqrt{(A_{mut} - A_{wt})^2 + (C_{mut} - C_{wt})^2 + (G_{mut} - G_{wt})^2 + (T_{mut} - T_{wt})^2},$$

where A_{mut} , C_{mut} , G_{mut} , and T_{mut} , respectively represent the frequency of A, C, G, and T bases in the mixed mutant pool, whereas A_{wt} , C_{wt} , G_{wt} , and T_{wt} respectively represent the frequency of A, C, G, and T bases in the mixed wild-type pool. During the analysis, there was a SNP site that differed between the two mixed genotypes. Additionally, the depth of the individual bases in the different mixed pools was statistically significant and the ED value for each site was calculated. The background noise was eliminated by applying the 5th power of the original ED as the correlation value (Hill et al., 2013). The distance method was then used to fit the ED value. The median + three standard deviations of the fitted values for all loci was set as the correlation threshold (Hill et al., 2013) to identify candidate regions. Using InDels for the association analysis, which was performed according to the method used for the SNP association analysis, the final candidate regions were identified by determining the intersecting regions corresponding to SNPs and InDels.

RESULTS

Analysis of the Genetic Diversity of an *in situ* Conserved Wild Rice Population

The *in situ* conservation site for wild rice in this study was located in Guiping county, Guangxi province, China (N 22°52', E 109°41') (Figure 1A). This site includes a typical natural *O. rufipogon* population. First, it is located at the center of wild rice genetic diversity in China. Second, this population is the biggest natural *O. rufipogon* population worldwide. This *in situ* conservation site, which was constructed in 2001, comprises 4.21 ha with 2.6 million *O. rufipogon* plants (Xu et al., 2020). An examination of plant architecture revealed the following four plant types: erect, semi-erect, tilted, and creeping (Figure 1B), which accounted for 1.9, 2.5, 4.4, and 91.2% of the whole population, respectively.

The genetic diversity of this population was determined by analyzing 184 individual plants. More than 80 SSR and InDel primer pairs stored in our laboratory were included in the analysis. Finally, 16 SSR markers and 24 InDel markers that were polymorphic and evenly disturbed on 12 chromosomes were selected for further study (Supplementary Table 1). The mean values for *Ae* (effective number of alleles), *He*, *Ho*, and *I* across all loci were 1.204, 0.852, 0.148, and 0.259, respectively (Table 1). The genetic relationships among samples were analyzed using STRUCTURE and the constructed UPGMA dendrogram. For the structural analysis, when the population number *K* was 5, the log likelihood $\ln[P(D)]$ was highest (Figure 1C). Accordingly, five groups were determined (Figure 1D). All erect and semi-erect individuals were classified in Group 2, whereas tilted individuals were classified in Groups 1 and 2 and creeping individuals were classified in all groups except for Group 2, although they were mainly in Groups 3–5. The UPGMA dendrogram revealed five main clusters, with similar structural results (Figure 1E).

Comparative Analysis of the Guiping Wild Rice and Cultivated Rice Genomes

Twenty *O. rufipogon* accessions in the *in situ* conservation site were selected for genome sequencing (i.e., three erect, four semi-erect, five tilted, and eight creeping accessions). After aligning the reads to the rice reference genome sequence, we identified 5,037,497 non-singleton SNPs. On the basis of the SNP data, the sequence diversity (π) of *O. rufipogon* was estimated to be about 0.003, which is higher than that of *O. sativa* (0.0024), *O. sativa* ssp. *indica* (0.0016), and *O. sativa* ssp. *japonica* (0.0006) (Huang et al., 2012). A principal component analysis (PCA) indicated that these accessions could be grouped according to their plant architecture (Figure 2A). We also investigated the population structure of 451 wild rice accessions that were included in an earlier study (Huang et al., 2012). Interestingly, the neighbor-joining tree showed that the creep and most tilt phenotypes of Guiping wild rice comprises Or-III plants, most of which originated in Guangxi province. While erect and semi-erect wild rice clustered in Or-I, which clustered closer to *indica* (Figure 2B).

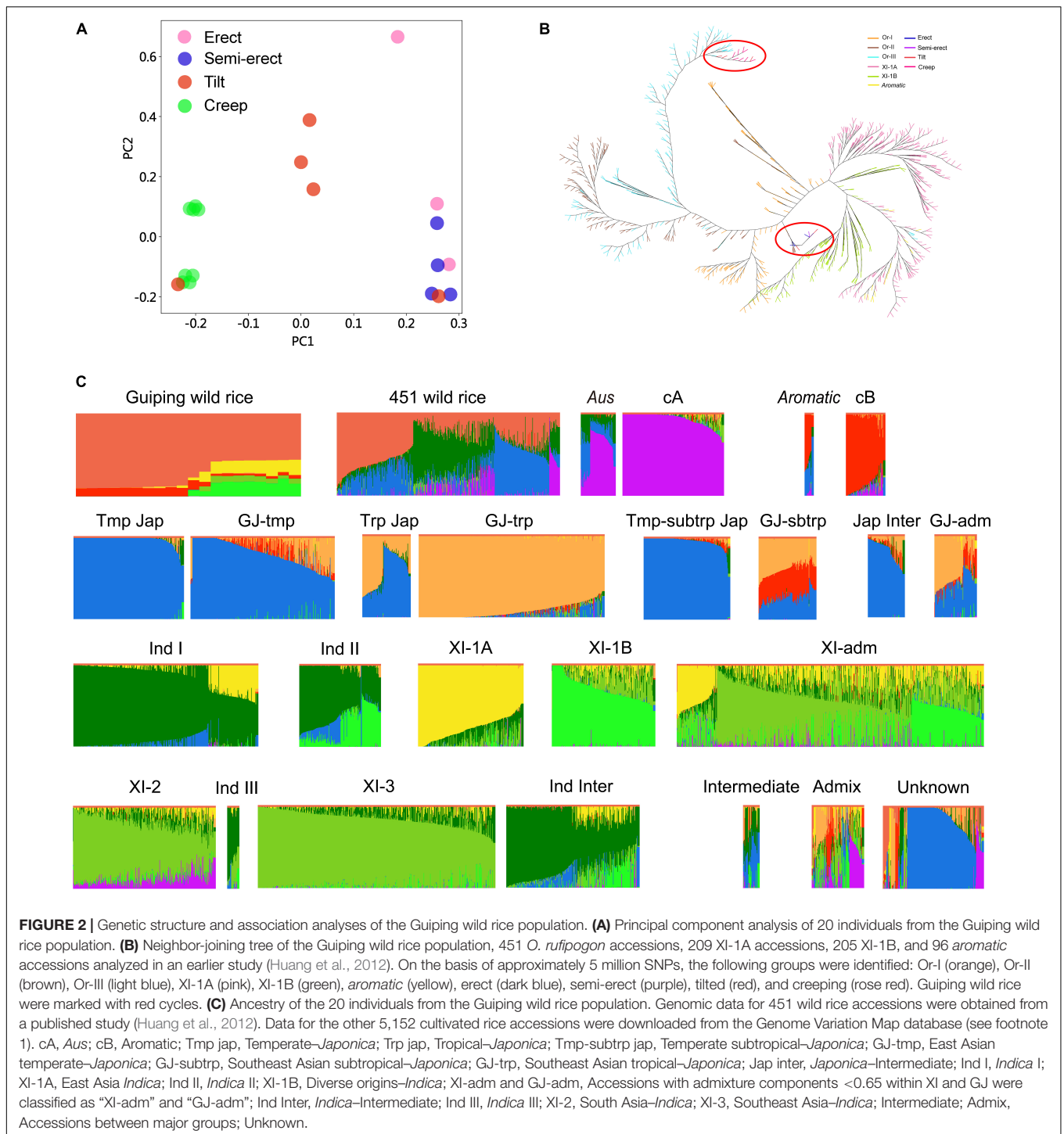
To compare our Guiping wild rice population with worldwide wild and cultivated rice populations, the whole-genome

sequencing data for 5,152 cultivated rice accessions (5K database) were downloaded from the Genome Variation Map database.¹ We compared our data with the 5K data and the published data for 451 wild rice accessions (Huang et al., 2012). Germplasm groups were defined by ADMIXTURE (Alexander et al., 2009), with the number of ancestral populations (*K*) ranging from 6 to 9 (Figure 2C and Supplementary Figure 1). The results indicated that most of the genetic components in the 20 Guiping wild rice accessions were from wild rice, but some

¹<https://ngdc.cncb.ac.cn/gvm/showIndividuals?type=2&orgId=2>

TABLE 1 | Genetic diversity-related parameters for different SSR and InDel loci.

Locus	Chromosome	He	Ho	Nei	Ae	I
Indel1-4	1	0.859	0.141	0.1406	1.1636	0.2691
Indel1-9	1	0.9892	0.0108	0.0108	1.0109	0.0338
RM10716	1	0.8727	0.1273	0.127	1.15	0.2428
Indel2-4	2	0.8498	0.1502	0.1498	1.1761	0.2825
Indel2-3	2	0.8205	0.1796	0.1791	1.2432	0.3066
RM12923	2	0.8518	0.1482	0.1478	1.1978	0.2559
RM13406	2	0.5024	0.4976	0.4962	1.985	0.6894
Indel3-7	3	0.8835	0.1165	0.1162	1.1371	0.225
Indel3-23	3	0.9784	0.0216	0.0215	1.022	0.06
RM14759	3	0.8408	0.1592	0.1588	1.1888	0.2954
RM15347	3	0.5982	0.4018	0.4007	1.6686	0.5903
Indel4-3	4	0.8683	0.1317	0.1313	1.1512	0.2553
Indel4-10	4	0.9784	0.0216	0.0215	1.022	0.06
RM16555	4	0.9053	0.0947	0.0945	1.1083	0.1915
RM5424	4	0.8057	0.1943	0.1938	1.2403	0.3438
Indel5-2	5	0.9784	0.0216	0.0215	1.022	0.06
Indel5-9	5	0.7972	0.2028	0.2022	1.2535	0.3551
RM18502	5	0.8683	0.1317	0.1313	1.1512	0.2553
Indel6-4	6	0.8498	0.1502	0.1498	1.1761	0.2825
Indel6-9	6	0.8748	0.1252	0.1249	1.1467	0.241
RM19725	6	0.9367	0.0633	0.0631	1.0673	0.1437
RM20111	6	0.9784	0.0216	0.0215	1.022	0.06
Indel7-5	7	0.8057	0.1943	0.1938	1.2403	0.3438
Indel7-9	7	0.8173	0.1827	0.1822	1.2493	0.3112
RM21236	7	0.8408	0.1592	0.1588	1.1888	0.2954
Indel8-3	8	0.9784	0.0216	0.0215	1.022	0.06
Indel8-8	8	0.947	0.053	0.0529	1.0558	0.1248
RM22959	8	0.9784	0.0216	0.0215	1.022	0.06
Indel9-2	9	0.8683	0.1317	0.1313	1.1512	0.2553
Indel9-8	9	0.9241	0.0759	0.0757	1.0873	0.1515
RM23842	9	0.6531	0.3469	0.3459	1.5547	0.5255
Indel10-6	10	0.7807	0.2193	0.2188	1.28	0.3768
Indel10-8	10	0.9626	0.0374	0.0373	1.0388	0.094
RM25375	10	0.6698	0.3302	0.3293	1.5197	0.5051
Indel11-4	11	0.7461	0.2539	0.2533	1.3924	0.3944
Indel11-9	11	0.8873	0.1127	0.1124	1.1267	0.2264
RM26319	11	0.8407	0.1593	0.1589	1.2266	0.2675
Indel12-3	12	0.8172	0.1828	0.1823	1.2297	0.3242
Indel12-9	12	0.9383	0.0618	0.0616	1.0688	0.1301
RM28107	12	0.7279	0.2721	0.2714	1.4008	0.4352
Mean		0.8518	0.1482	0.1478	1.204	0.2595



were from Aromatic, XI-1A, and XI-1B ($K = 9$, **Figure 2C**). They were divided into the following nine groups: *Indica*, *Japonica*, Tropical-Japonica, Temperate-Japonica, *O. rufipogon*, *Aus*, *Aromatic*, Intermediate-type, and Other. An examination of the Guiping wild rice accessions with erect or semi-erect phenotypes revealed their genetic components were mainly from wild rice (about 60%), XI-1B (about 20%), and XI-1A (about 15%), but

a small percentage was from Aromatic. In contrast, the genetic components of the accessions with a tilted phenotype were primarily from wild rice (about 80%) and Aromatic (about 10%), but a small proportion was from XI-1A. The genetic components of the accessions with a creeping phenotype were mostly from wild rice (about 90%), with approximately 10% from Aromatic. The ABBA results also indicated that the 20 Guiping wild rice

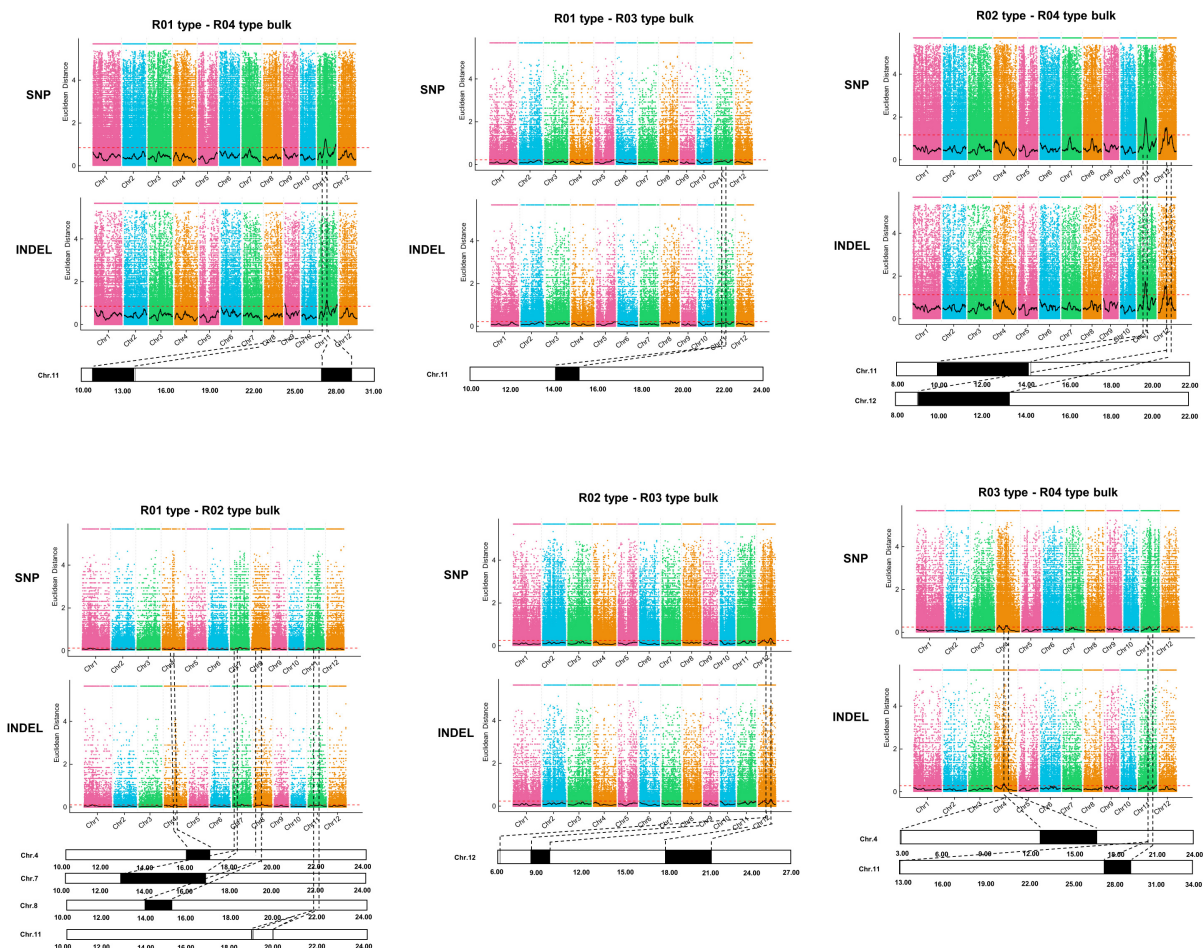


FIGURE 3 | Identification of genomic regions controlling plant architecture in the Guiping wild rice population according to the PCAMP data. Four plant architecture types: R01-erect, R02-semi-erect, R03-tilted, and R04-creeping. Thirty individuals were selected for each type to construct a mixed DNA pool. A bulked segregant analysis was performed for each type. The Euclidean distance (ED) value of SNPs and InDels was calculated to identify the candidate regions for each comparison. The overlapping physical positions on the same chromosome for each comparison are in black.

TABLE 2 | Final genomic candidate regions related to the plant architecture in the Guiping wild rice population.

Chromosome	Genomic candidate regions (Mb)	Known genes
4	16.09–17.18	DHD1
8	14.27–15.43	
11	11.11–11.28	
11	27.89–29.01	
12	9.62–10.41	

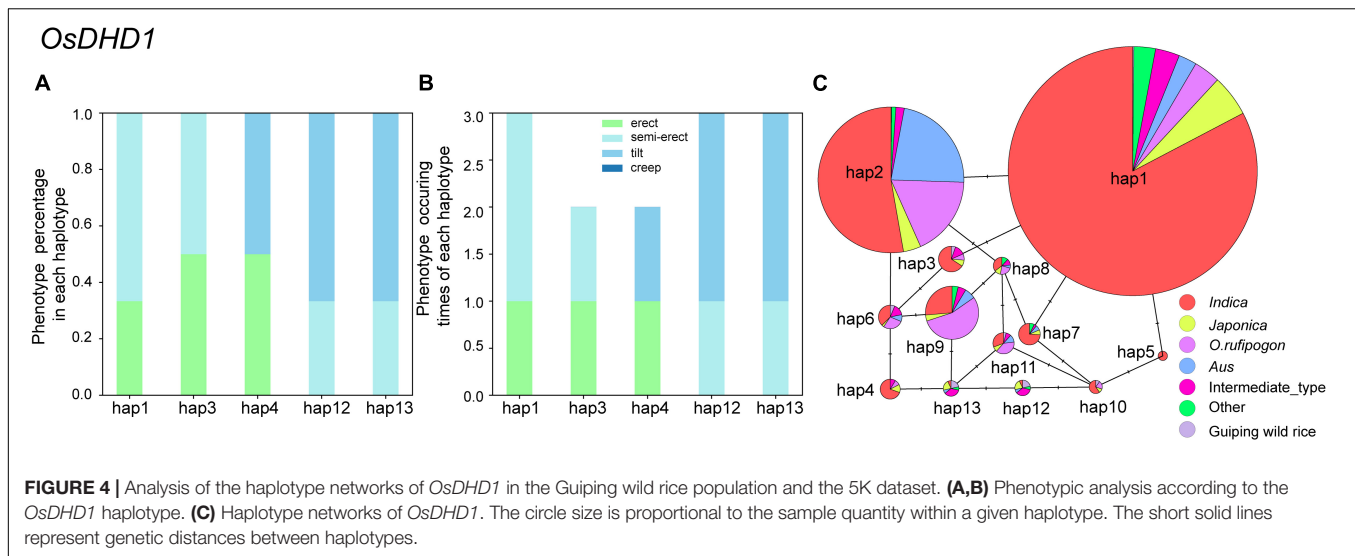
accessions had genetic components from Aromatic, XI-1A, and XI-1B (Supplementary Table 2).

Identification of Main-Effect Variations in the Plant Architecture of Guiping Wild Rice

A PCAMP protocol (Yang et al., 2019) was used to identify candidate genomic regions related to plant architecture in

the wild rice population. The whole population was divided into the following four subpopulations that differed in terms of plant architecture: R01-erect, R02-semi-erect, R03-tilted, and R04-creeping. Genomic DNA was extracted from 30 phenotypically identical individuals from each subpopulation and then combined as one bulk for the pair-wise comparison. The R01, R02, R03, and R04 mixed DNA was sequenced using the Illumina HiSeq X 10 high-throughput sequencing platform. After filtering the raw data, 85.18 Gb remained, of which R01, R02, R03, and R04 accounted for 13.81, 17.29, 20.75, and 34.33 Gb, respectively. The average sequencing depth was 48.5-fold, covering 97.76% of the whole genome.

To identify candidate genomic regions responsible for the plant architecture of the Guiping wild rice population, we compared the SNPs and InDels between DNA pools. We used ED-based methods involving SNPs and InDels to identify candidate regions. The overlapping physical positions on the same chromosome were selected in each comparison (Figure 3). The intersecting regions (Table 2) were selected as the final



candidate regions that were significantly associated with plant architecture-related genes in the wild rice population. A *DHD1* gene on chromosome 11 had the highest ED value in the R01–R04 comparison. In a recent study, *DHD1* was revealed to regulate the rice heading date, while also influencing the plant architecture (Zhang et al., 2019). No other reported plant architecture-related genes were detected in these overlapping regions. Hence, *DHD1* might be the candidate gene mediating the variations in the plant architecture in the Guiping wild rice population. To confirm this possibility, we analyzed the following 10 well-documented plant architecture-related genes: *OsPROG1*, *OsLAZY1*, *OsMOC1*, *OsTAC1*, *OsLIC1*, *OsDLT*, *OsMOC3*, *OsTB1*, *OsTAD1*, and *OsDHD1* (Supplementary Table 3). Using data for our 20 Guiping wild rice accessions and 451 *O. rufipogon* accessions as well as the data in the 5K database, the genic and 1-kb flanking regions of these 10 genes were used for a variant analysis. Their haplotype networks were constructed according to the minimum spanning method implemented in POPART. Notably, Only *DHD1* had a unique haplotype related to the erect/semi-erect phenotype in the Guiping wild rice population. Other genes, including *OsPROG1*, *OsLAZY1*, and *OsMOC1* (Li et al., 2003; Li et al., 2007; Tan et al., 2008), which significantly influence rice plant architecture, had no erect phenotype-related haplotypes in the Guiping wild rice population (Supplementary Figures 2, 3). Thirty haplotypes were detected for *OsDHD1*, of which haplotypes 1, 3, and 4 were associated with the erect or semi-erect phenotype (Figure 4). The cultivated rice accessions with these haplotypes were mainly *indica* accessions belonging to Subgroup XI-A from East Asian countries, among them more than 52% (1408/2962) accessions were from China.

DISCUSSION

Oryza rufipogon is an important germplasm resource for food security, but it is seriously endangered worldwide. Almost

all studies on the *in situ* conservation of wild rice have been conducted in China. Most of these studies focused on the related policy, techniques, and methods for establishing *in situ* conservation sites. Thus, there are relatively few reports describing genetic analyses of the introgression involving other populations, including cultivated rice and other wild rice accessions.

In this study, a typical *in situ* conserved Chinese wild rice population in Guiping county of Guangxi province was investigated. The genetic diversity was first elucidated on the basis of 40 molecular markers (Supplementary Table 1). Compared with the genetic diversity detected in previous studies, the genetic diversity of the *in situ* conserved Guiping population was relatively stable. For example, Zhou et al. (2003) detected 12 populations in Guangxi, Gao (2004) analyzed six wild rice populations across China, and Yu et al. (2004) investigated the genetic diversity of *O. rufipogon* in many regions in Guangxi. These studies included the Guiping population, but samples were derived from germplasm (seed) banks or gardens (i.e., *ex situ* conservation sites). The genetic diversity indicators *Ho* and *I* in the current study were consistent with the corresponding values in these earlier studies. Considering the materials from *ex situ* conservation sites were collected from their original habitats 10 or 20 years ago, the *in situ* conservation site in Guiping was suitable for maintaining wild rice genetic diversity. Moreover, the sequence diversity (π) determined according to the SNP data (about 0.003) was similar to that of wild rice and much higher than that of cultivated rice in an earlier study (Huang et al., 2012). Accordingly, at the center of wild rice genetic diversity in China, the Guiping wild rice population is highly genetically diverse.

The central goal of the *in situ* conservation of wild rice is to prevent human activities from affecting genetic diversity and cultivated rice gene flow. We re-sequenced 20 individuals from this conservation site that varied in terms of plant architecture (Figure 1B). The PCA results implied that these 20 individuals could be grouped according to their plant

architecture (**Figure 2A**). Additionally, the analysis of genetic structures involving SSR/InDel markers indicated that this population can be divided into five subpopulations (**Figure 1D**). These results revealed that the *in situ* conserved wild rice population in Guiping has a special plant architecture-related population structure. On the basis of published data (Huang et al., 2012), all of the samples collected from Guiping in this study were identified as Or-III plants (**Figure 2B**), indicative of a strong correlation with geographic distribution. The introgression between wild rice and XI-1A or XI-1B was also confirmed in an earlier investigation (Huang et al., 2012). However, there are no reports suggesting that wild rice has genetic components from aromatic rice. In the current study, we detected genes in the Guiping wild rice genome that were introgressed from aromatic rice populations. Moreover, ABBA-BABA data also suggested that the Guiping wild rice genome includes genes from *indica* rice (**Supplementary Table 2**).

The Guiping wild rice population, which is the largest *in situ* conserved wild rice population, grows in the central Pearl River region, which is where rice was first domesticated (Huang et al., 2012). There were no plants with an erect phenotype when the conservation site was established 20 years ago. To clarify how variations in plant architecture developed, we identified the main-effect variants using a PCAMP approach and constructed haplotype networks for 10 plant architecture-related genes. Introgressions were revealed only for *OsDHD1*. The PCAMP results also suggested that *OsDHD1* may be a candidate gene associated with plant type variations. We also detected variations in major plant architecture-related genes, such as *PROG1* and *LAZY1*, in the wild rice population, but they did not affect plant architecture. An examination of the cultivated rice accessions that were the source of the introgressed genes in Guiping wild rice revealed a lack of accessions in Guiping with *OsDHD1* haplotypes 1, 3, or 4, but a few cultivated rice varieties or landraces native to Guangxi province with these haplotypes were identified.

Because only growing closely related *Oryza* species may result in wild rice pollen infiltration, we speculated that the plant architecture variations may have been caused by the introgression of genes from Guiping landraces, but also by the infiltration of pollen from major commercially cultivated rice varieties. A good plant structure may positively affect the planting density and photosynthetic efficiency, while also providing plants with a competitive advantage. Therefore, variations in plant architecture may be the result of natural genomic modifications that enable adaptations to environmental conditions. The Guiping *in situ* wild rice conservation site has metal fences to keep out livestock and is surrounded by many fields containing cultivated rice varieties. Accordingly, we strongly recommend that the *in situ* conservation site should be surrounded by a fence that can appropriately isolate the area and prevent the infiltration of alien pollen. Furthermore, in future related genomic investigations, additional individuals should be collected for *ex situ* conservation, identification, and evaluation to conserve the genetic integrity of the population and to

identify elite genes in the population relevant for improving rice production.

DATA AVAILABILITY STATEMENT

The data presented in this study are deposited in NCBI repository, accession number: SRR19425523–SRR19425542.

AUTHOR CONTRIBUTIONS

ZY and YZ performed most of the experiments, including phenotyping, data analysis, and haplotype analysis. YZ, MX, and XW participated in the genomic comparison. JH and ZX participated in some of the phenotyping and genetic diversity analysis. FL, YW, YN, JG, DL, ZL, ZH, YL, and XZ participated in the field investigation and logistic work. QY supervised the study. HH and WQ designed the study and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Key Research and Development Program of China (2021YFD1200501), project of Sanya Yazhou Bay Science and Technology City (Grant No. SKJC-2020-02-001), and Hainan Yazhou Bay Seed Laboratory (project of B21HJ0215).

ACKNOWLEDGMENTS

We thank Fu-Guang Deng, Zong-Yong Meng, Yun-Shao Meng, the staff of Bureau of Agriculture and Rural Affairs, Guiping County Government, and Guangxi Zhuang Autonomous Region of China, for their assistance to the material collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.921349/full#supplementary-material>

Supplementary Figure 1 | Genetic structure and association analysis of Guiping wild rice population ($K=6-8$). The represented subgroup as same as **Figure 2**.

Supplementary Figure 2 | Haplotype networks analysis of plant architecture related genes *OsPROG1*, *OsLAZY1*, and *OsMOC1* in Guiping wild rice population and 5K dataset. For each gene, left two figures represented phenotypic analysis according to the haplotype, and right figure represented the Haplotype networks.

Supplementary Figure 3 | Haplotype networks analysis of other plant architecture related genes *OsTAC1*, *OsLIC1*, *OsDLT*, *OsMOC3*, *OsTB1*, and *OsTAD1* in Guiping wild rice population and 5K dataset. For each gene, left two figures represented phenotypic analysis according to the haplotype, and right figure represented the Haplotype networks.

REFERENCES

- Alexander, D. H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* 19, 1655–1664. doi: 10.1101/gr.094052.109
- Cingolani, P., Platts, A., Wangle, L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Civan, P., Craig, H., Cox, C. J., and Brown, T. A. (2015). Three geographically separate domestications of Asian rice. *Nat. Plants* 1:15164. doi: 10.1038/nplants.2015.164
- Edwards, K., Johnstone, C., and Thompson, C. (1991). A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic. Acids. Res.* 19:1349.
- Falush, D. M. S., and Pritchard, J. K. (2003). Inference of population structure using multi-locus genotype data: linked loci and correlated allele frequencies. *Genetics* 164, 1567–1587. doi: 10.1093/genetics/164.4.1567
- Fuller, D. Q., Sato, Y. I., Castillo, C., Qin, L., Weisskopf, A. R., Kingwell, B. J., et al. (2010). Consilience of genetics and archaeo-botany in the entangled history of rice. *Archaeol. Anthropol. Sci.* 2, 115–131.
- Gao, L. (2004). Population structure and conservation genetics of wild rice *Oryza rufipogon* (Poaceae): a region-wide perspective from microsatellite variation. *Mol. Ecol.* 13, 1009–1024. doi: 10.1111/j.1365-294X.2004.02108.x
- Gao, L., Zhang, S., Zhou, Y., Ge, S., and Hong, D. A. (1996). Survey of the current status of wild rice in China. *China Biodivers.* 48, 160–166.
- Hill, J. T., Demarest, B. L., Bisgrove, B. W., Gorski, B., Su, Y., and Yost, H. (2013). MMAPP: mutation mapping analysis pipeline for pooled RNA-seq. *Genome Res.* 23, 687–697. doi: 10.1101/gr.146936.112
- Huang, X., Kurata, N., Wei, X., Wang, Z., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532
- Jakobsson, M., and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 1801–1806. doi: 10.1093/bioinformatics/btm233
- Jin, J., Huang, W., Gao, J., Yang, J., Shi, M., Zhu, M., et al. (2008). Genetic control of rice plant architecture under domestication. *Nat. Genet.* 40, 1365–1369. doi: 10.1038/ng.247
- Lee, T. H., Amber, N. H., Hannah, R., Scott, A. J., Soraya, C. M., Leal-Bertioli, M. T., et al. (2019). Breeding crops to feed 10 billion. *Nat. Biotechnol.* 37, 744–754. doi: 10.1038/s41587-019-0152-9
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, P. J., Wang, Y. H., Qian, Q., Fu, Z. M., 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
- Li, X. Y., Qian, Q., Fu, Z. M., Wang, Y. H., Xiong, G., Zeng, D., et al. (2003). Control of tillering in rice. *Nature* 422, 618–621.
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Nei, M. (1972). Genetic distances between populations. *Am. Nat.* 106, 283–292.
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89, 583–590.
- Oka, H. I. (1988). *Origin of Cultivated Rice*. New York: Elsevier.
- Ramakrishnan, M., Antony Cesar, S., Duraipandiyan, V., Al-Dhabi, N. A., and Ignacimuthu, S. (2016). Using molecular markers to assess the genetic diversity and population structure of finger millet (*Eleusine coracana* (L.) Gaertn.) from various geographical regions. *Genet. Resour. Crop Evol.* 63, 361–376.
- Tan, L., 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
- Wang, H., Vieira, F. G., Crawford, J. E., Chu, C., and Nielsen, R. (2017). Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. *Genom Res.* 27, 1029–1038. doi: 10.1101/gr.204800.116
- Xu, Z. J., Wang, J. L., Zheng, X. M., Fang, Z. L., Tang, C. F., Wang, X. H., et al. (2020). Investigation, collection and protection of wild rice germplasm resources in China. *J. Plant Genet. Resour.* 21, 1337–1343.
- Yang, X., Xia, X., Zhang, Z., Nong, B., Zeng, Y., Wu, Y., et al. (2019). Identification of anthocyanin biosynthesis genes in rice pericarp using PCAMP. *Plant Biotechnol. J.* 17, 1700–1702. doi: 10.1111/pbi.13133
- Yang, Z., Xu, Z., Yang, Q., and Qiao, W. (2022). Conservation and utilization of genetic resources of wild rice in China. *Rice Sci.* 29, 216–224.
- Yeh, F. C., Re, Y., and Boyle, T. (1998). POPGENE version 1.31. Microsoft windows-based freeware for population genetic analysis. *Univ. Alberta Cent. Int. For. Res.* 11–23.
- Yu, P., Li, Z., Zhang, H., Li, D., Wang, M., and Sun, J. (2004). Phenotypic traits and ssr diversity of common wild rice (*Oryza rufipogon* Griff.) in Guangxi. *Acta Genet.* 31, 934–940.
- Zhang, H., Zhu, S., Liu, T., Wang, C., Cheng, Z., Zhang, X., et al. (2019). DELAYED HEADING DATE1 interacts with OsHAP5C/D, delays flowering time and enhances yield in rice. *Plant Biotechnol. J.* 17, 531–539. doi: 10.1111/pbi.12996
- Zhou, H. F., Xie, Z. W., and Ge, S. (2003). Microsatellite analysis of genetic diversity and population genetic structure of a wild rice (*Oryza rufipogon*) in China. *Theor. Appl. Genet.* 107, 332–339. doi: 10.1007/s00122-003-1251-y

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Yang, Zhang, Xing, Wang, Xu, Huang, Wang, Li, Nie, Ge, Lou, Liu, Han, Liang, Zheng, Yang, He and Qiao. 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(s) 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.



Germplasm Sources, Genetic Richness, and Population Differentiation of Modern Chinese Soybean Cultivars Based on Pedigree Integrated With Genomic-Marker Analysis

OPEN ACCESS

Edited by:

Xia Xin,
Institute of Crop Sciences
(CAS), China

Reviewed by:

Donghe Xu,
Japan International Research Center
for Agricultural Sciences
(JIRCAS), Japan
Hamid Khazaei,
World Vegetable Center, Taiwan

*Correspondence:

Jinming Zhao
jzm3000@126.com
Junyi Gai
sri@njau.edu.cn

†Present address:

Xiaoyan Zhang,
Institute of Crop Germplasm
Resources, Shandong Academy of
Agricultural Sciences, Jinan, China

Specialty section:

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

Received: 17 May 2022

Accepted: 15 June 2022

Published: 11 July 2022

Citation:

Li C, Wang W, Pan Y, Liu F, He J,
Liu C, Cao J, Zhang X, Zhao J and
Gai J (2022) Germplasm Sources,
Genetic Richness, and Population
Differentiation of Modern Chinese
Soybean Cultivars Based on Pedigree
Integrated With Genomic-Marker
Analysis. *Front. Plant Sci.* 13:945839.
doi: 10.3389/fpls.2022.945839

Chunyan Li^{1,2}, Wubin Wang¹, Yongpeng Pan¹, Fangdong Liu¹, Jianbo He¹,
Chuanxiang Liu², Jiqui Cao², Xiaoyan Zhang^{2†}, Jinming Zhao^{1*} and Junyi Gai^{1*}

¹ Soybean Research Institute & MARA National Center for Soybean Improvement & MARA Key Laboratory of Biology and Genetic Improvement of Soybean (General) & State Key Laboratory for Crop Genetics and Germplasm Enhancement & Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing Agricultural University, Nanjing, China, ² MARA Key Laboratory of Germplasm Enhancement and Breeding Technology of Soybean, Jiaxiang, China

Soybean is a native crop in China for $\approx 5,000$ years. The 560 cultivars released in 2006–2015, commercialized with seeds available publicly, were collected (designated modern Chinese soybean cultivars, MCSCs), as a part of 2,371 ones released during ~ 100 years' breeding history. The MCSCs with their parental pedigrees were gathered, including 279, 155, and 126 cultivars from Northeast and Northwest China (NNC), Huang-Huai-Hai Valleys (HHH), and Southern China (SC), respectively. The MCSCs were tested in the field, genotyped with sequencing, and analyzed for their germplasm sources, genetic richness, and population differentiation based on pedigree integrated with genomic-marker analysis. The main results were as follows: (i) The MCSCs covering 12 of the global 13 MGs (maturity groups) showing different ecoregions with different cropping systems caused their different MG constitutions. (ii) Parental pedigree analysis showed 718 immediate parents and 604 terminal ancestors involved in MCSCs, from which 41 core-terminal ancestors were identified. (iii) NNC was richer in allele number and specific present/deficient alleles, and genetically distant from HHH and SC. (iv) The geographic grouping of MCSCs was partially consistent with marker-based clustering, indicating multiple genetic backgrounds in three eco-subpopulations. (v) Eleven major core-terminal ancestor-derived families were identified, including four derived from ancestors in NNC, four from HHH, and three from SC, containing 463 (82.68%) MCSCs with some cross-distribution among ecoregions. (vi) CGS (coefficient of genetic similarity) calculated from genomic markers showed more precision than COP (coefficient of parentage) using pedigree information in evaluating genetic relationship/differentiation. Overall, through pedigree and genomic-marker analyses, the germplasm constitutions of the three eco-subpopulations were relatively self-sufficient, and germplasm exchange is seriously required for further improvement.

Keywords: modern Chinese soybean cultivar population (MCSCs/MCSCP), germplasm, pedigree, core-terminal ancestor, population differentiation, coefficient of parentage (COP), coefficient of genetic similarity (CGS)

INTRODUCTION

Soybean (*Glycine max* (L.) Merr.), originated in ancient China $\approx 5,000$ years ago (Zhao and Gai, 2004), and has disseminated all around the world as an economic crop rich in protein ($\sim 40\%$) and oil ($\sim 20\%$). In China, soybeans have been widely planted, covering almost all the arable area, all the maturity groups except MG X, and different kinds of cropping systems (Gai and Wang, 2001). Especially, during the recent decades, soybean has expanded to further higher latitudes in mono-cropping and to further early seasons in multi-cropping systems. Now, the soybean production areas in China can be roughly grouped into three major ecoregions, i.e., the Northeast and Northwest China northern spring soybean cultivation region (NNC, Northwest region was newly added), the Huang-Huai-Hai Valleys spring-summer double-cropping soybean cultivation region (HHH), and the Southern China spring-summer-autumn multi-cropping soybean cultivation region (SC).

Throughout history, ancient farmers developed soybean varieties for their own production, and a large number of farmers' varieties or landraces formed and accumulated. Along with the new landraces formed, some old ones retired. Fortunately, modern cool storage techniques saved the remained landraces. Based on the landraces, modern soybean breeders developed and released soybean cultivars scientifically, which was started in the early 20 century in China (Gai et al., 2015). In 1923–2015, altogether, 2,371 cultivars have been released in the three soybean eco-regions in China.

An effective and efficient plant breeding depends on the source of raw materials or the germplasm reservoir (gene pool) (Gai et al., 2015). In the early time, the soybean germplasm mainly included landraces and annual wild relatives (*Glycine soja* Sieb. & Zucc). With the continuous release of new cultivars used currently then turning as breeding materials for the next round of breeding, the released cultivars as well as their derived specific materials through modern technologies become a growing source of germplasm resources/reservoir, especially accompanied by the joining of various specific introductions during the breeding history. In spite of the importance of released cultivars as growing germplasm components in plant breeding, it is especially necessary to explore the phenotypic and genotypic characteristics of the released cultivars for further genetic improvement. Based on it, the germplasm/gene backgrounds and potentials, germplasm/gene deficiencies and introduction requirements, as well as new technique preparations for achieving further breeding targets may be prepared and designed in advance.

In exploring the phenotypic and genotypic characteristics of the released cultivars as the germplasm source, it has experienced three stages. The initial stage was to collect and catalog the released cultivars in addition to having the landraces collected already, with their geographical site, genetic sources, and morphological and agronomical traits evaluated and documented. In the second stage, their genetic utility was evaluated through parental pedigree analysis, for which the coefficient of parentage (COP) between/among cultivars was estimated to indicate their genetic relationship. With the rapid development of molecular biology, the third stage started by using molecular marker technology. By which the

genetic relationship between/among varieties can be evaluated with a coefficient of genetic similarity (CGS) and the genetic constitution of each trait, i.e., QTL-allele or gene-allele constitutions in the population can be explored through the QTL/gene mapping procedure (Zhang et al., 2015; Meng et al., 2016; Li et al., 2017; Fu et al., 2020; Liu et al., 2021a). Based on it, the QTLs-alleles or genes-alleles can be traced in the population. This kind of study started from released cultivars and now has been extended to landraces as well as wild soybean germplasm resources. As indicated by Liu et al. (2021b), the world's 90% of soybean production is related to the soybean germplasm from Northeast China, understanding the genetic constitution of this population, in fact, is a short way to know the genetic background of the world major soybean sources.

Gai et al. (2015) summarized the pedigree analysis results as well as their germplasm bases of the 1,300 soybean cultivars released from 1923 to 2005 in China. Their work was at the second stage basically, regarding the genetic relationship through parental pedigree analysis. They showed all the parental pedigrees of 1,300 soybean cultivars released from 1923 to 2005 and traced their terminal ancestors to reveal the major germplasm sources. The cultivar improvement developed quickly after 2005 in China, there were 1,071 cultivars released at the state-/province-authorized level from 2006 to 2015. Among them, 560 were commercialized with seeds available publicly that have been collected and designated as modern Chinese soybean cultivar population (MCSCP/MCSCs). As indicated above, these recently released cultivars are historical accumulations of previous breeding efforts from 1923 through 2015. Because each round of the breeding cycle laid on the previous cultivar generations with the superior genes-alleles accumulated consecutively in the recent population. Based on this understanding, exploring the phenotypic and genotypic characteristics is in fact to understand the core germplasm of Chinese current soybean cultivars.

The present study aimed at exploring the major germplasm sources in the whole country as well as in the three ecoregions through pedigree analysis, summarizing the core ancestors of the MCSCP, and exploring the genetic richness, genetic similarity/diversity, and genetic differentiation of the MCSCP, as well as in three ecoregions through genomic marker analysis. Then, the study worked on exploring the genetic differentiation among the three ecoregions to see whether the geographic differentiation coincided with the genetic differentiation and then exploring the genetic differentiation of the MCSCP in terms of major core-terminal ancestor-derived families to see how the major terminal ancestors played their important roles in Chinese modern soybean cultivars. In addition, the COP values were compared to CGS values for their utilization in evaluating the genetic similarity/diversity among the released cultivars.

MATERIALS AND METHODS

The Collection of the Modern Chinese Soybean Cultivars Released From 2006 to 2015

A total of 560 soybean cultivars officially released from 2006 to 2015 were collected from the public and private breeders,

forming the MCSCP, including 279 from NNC, 155 from HHH, and 126 from SC (**Supplementary Figure 1A**). The 560 cultivars came from 26 provinces, including 122 state-authorized and 438 province-authorized cultivars (**Supplementary Table 1**). The passport data for all 560 cultivars can be referred to in **Supplementary Table 2**. The seeds of the MCSCs were collected directly from the respective breeders/institutions, while their parental pedigrees were mainly collected from the breeders as well as publications.

Field Experiments

In 2018, the 560 cultivars were tested in one-row plots, with a plot length of 1 m, and a row space of 0.4 m, in a randomized blocks design with two replications at the Dangtu experimental station of Nanjing Agricultural University in Anhui province, China, and three replications at Jiaxiang experimental station of the Key Laboratory of Germplasm Enhancement and Breeding Technology of Soybean in Shandong province, China. The sowing date was June 26 at Dangtu and June 23 at Jiaxiang. The evaluated traits were: beginning bloom (R1) and full maturity (R8) according to Fehr and Caviness (1977), days to flowering and days to maturity calculated as the days from sowing to R1 and R8, respectively, main stem node number counted at maturity with the cotyledonary node as zero up to the top, 100-seed weight evaluated three times per plot, and protein content and oil content evaluated per plot base using FOSS InfratecTM1241 (Hoganas, Sweden).

Statistical Analysis

PROC MEANS in SAS/STAT 9.4 (SAS Institute Inc., Cary, NC, USA) was used to calculate the descriptive statistics (mean, minimum, and maximum) of days to flowering, days to maturity, main stem node number, 100-seed weight, seed protein content, and seed oil content. PROC GLM was used for the analysis of variance (ANOVA), the linear model is

$$y_{ijk} = \mu + t_i + b_{j(i)} + g_k + (gt)_{ik} + \varepsilon_{ijk}$$

where μ is the population mean; t_i is the effect of i th environment; $b_{j(i)}$ is the effect of j th block within the i th environment; g_k is the effect of k th genotype and $(gt)_{ik}$ is the genotype-by-environment interaction effect; ε_{ijk} is the random error and $\varepsilon_{ijk} \sim N(0, \sigma^2)$. The variance components were estimated according to the expected mean squares in ANOVA. The genetic coefficient of variation was calculated as $GCV = \sigma_g/\mu \times 100\%$, where σ_g is the square root of genotypic variance.

The maturity group of the 560 soybean cultivars was determined with the growth period record obtained from the field experiments according to the standards in the literature (Fu et al., 2016; Tang et al., 2016; Liu et al., 2017; Song et al., 2019).

Parental Pedigree, Coefficient of Parentage, and Core-Terminal Ancestor-Derived Family

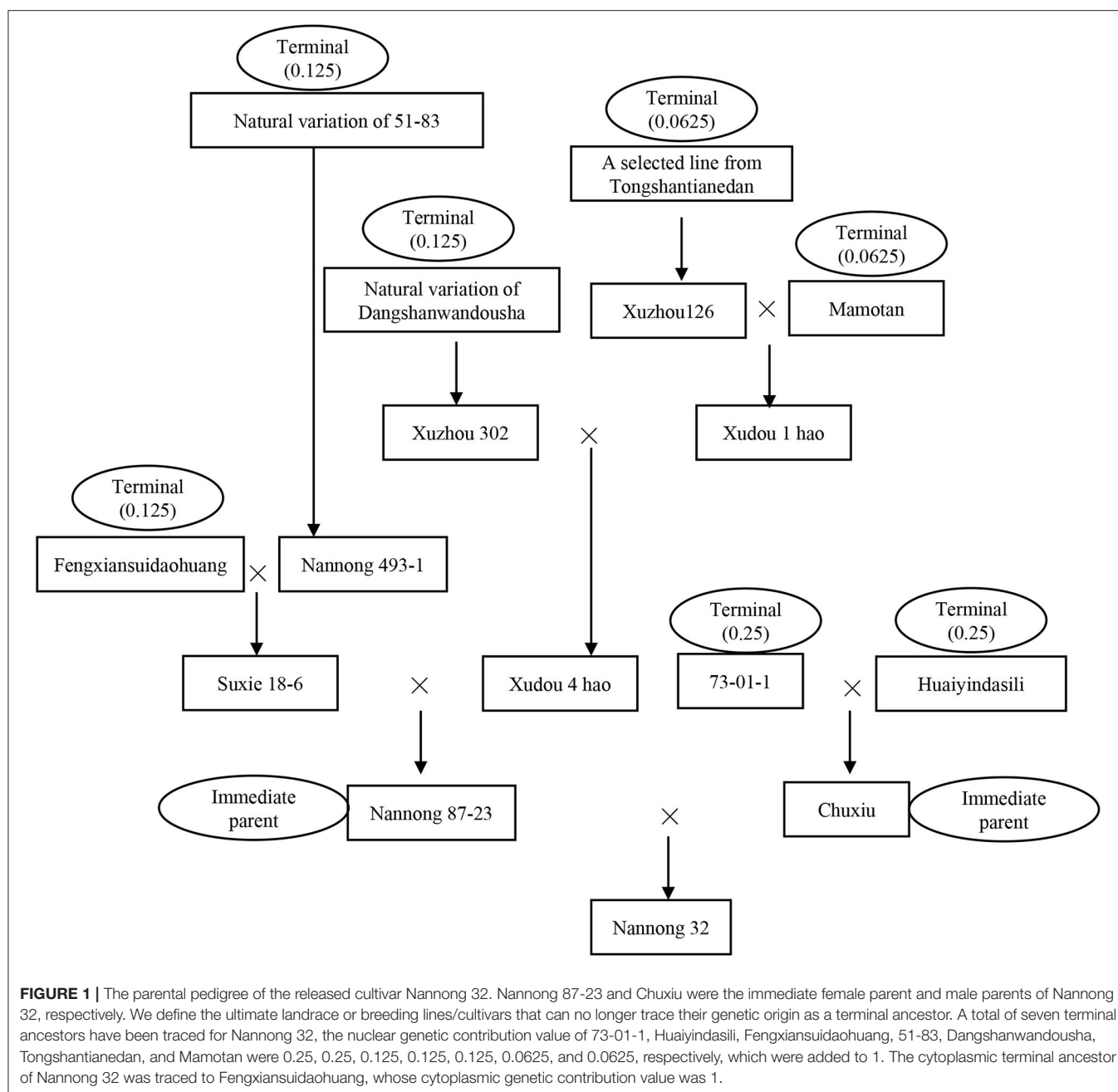
The method of parental pedigree analysis in this study was the same as that indicated by Gai et al. (1998) and Xiong et al.

(2008). In pure line selection or mutation breeding, the nuclear genetic contribution value of the source cultivar/variety/line to the pure line or mutant is recorded as 1. In hybridization breeding, the nuclear genetic contribution values of both the female and male parents to the hybrids are 1/2 or 0.5, the same proportion for the parents of the last hybridization, and up to terminal parents/ancestors, in this way, the total nuclear contribution of all parental ancestors to the derived cultivar must be 1. As for the cytoplasmic contribution, the terminal female ancestor of the immediate parent of the last round of hybridization contributes 1, while the others are zero. These roles are used in all kinds of crossing-types, such as the example of Nannong 32 in **Figure 1**, where the nuclear contribution values of the seven terminal ancestors in different crossing cycles to Nannong 32 were marked and added to 1. While the immediate cytoplasm parent Nannong 87-23 provided cytoplasm to Nannong 32, which can be traced to the terminal cytoplasm parent, a landrace of Fengxiansuidaohuang, to Nannong 32, so the cytoplasm contribution value of Fengxiansuidaohuang is 1.

The coefficient of parentage (COP) was calculated referring to Cox et al. (1985) and Cui et al. (2000). The basic rules are as the following: (i) A cultivar derived from hybridization obtained 50% of its germplasm from each parent. (ii) All terminal ancestors, parents, and derived cultivars or breeding lines were homozygous and homogeneous. (iii) The COP value of all terminal ancestors is 0. (iv) The COP value between the cultivar and its selected line and between a naturally occurred or mutagen-induced mutation and its antecedent are 0.75. (v) The COP value between the cultivars and itself is 1. (vi) The calculation formula of COP value between cultivars containing some of the same parents, is $COP_{SD} = \Sigma [(1/2)^n]$, where COP represents the coefficient of parentage between cultivar S and cultivar D, n represents the number of generations from the common parents to cultivar S and cultivar D. The population COP matrix was established based on all COP values among cultivars.

According to the geographic origin of the terminal ancestors of MCSCs, all the terminal ancestors were grouped into four sets NNC, HHH, SC, and foreign introduction. Those terminal ancestors of each set with three or more breeding cycles were picked up as candidate terminal ancestors, from which, those with two or more of the four indicators higher than the average of the candidate terminal ancestors were nominated as population-level core-terminal ancestors of the MCSCP. The four indicators were nuclear genetic contribution value, cytoplasmic genetic contribution value, number of derived cultivars, and the number of breeding cycles involved. In the core-terminal ancestors of the MCSCP, those involved in an eco-region are also the eco-regional core-terminal ancestors.

Based on the number of derived cultivars, the nuclear genetic contribution, and the cytoplasmic contribution of the core-terminal ancestors, 4, 4, and 3 major core-terminal ancestors with their derived cultivars were recognized as the major core-terminal ancestor-derived families in NNC, HHH, and SC, respectively.



Genotyping of the MCSCP and Assembly of Multi-Haplotype Genomic Markers

The whole-genome sequencing (WGS) was performed at Annoroad Ltd., Beijing, China. The total DNA samples of the 560 released cultivars were extracted from their young leaves using the CTAB method (Murray and Thompson, 1980). The sequences were obtained using Illumina HiSeq 2000 instrument through the multiplexed shotgun genotyping method (Andolfatto et al., 2011) with DNA fragments between 400 and 600 bp, generating 8.42 billion paired-end reads of 150 bp (1262.66 Gb of sequence), $\approx 2.30 \times$ in depth and 73.7%

coverage. All sequence reads were aligned against the reference genome Wm82.v2 (Schmutz et al., 2010) using BWA (Li, 2013). The GATK software (McKenna et al., 2010) was applied for population SNP calling. The SNPs of the 560 cultivars were polymorphic with a missing rate $\leq 20\%$, heterozygous rate $\leq 10\%$, and minor allele frequency (MAF) ≥ 0.01 . The Beagle software (Browning et al., 2018) was used for genotyping SNP imputation after heterozygous alleles were turned into missing alleles.

The whole-genome sequence was divided at a certain LD value ($D' > 0.7$) into SNP linkage disequilibrium blocks (SNPLDBs) as

genomic markers with their multiple haplotypes/alleles using the RTM-GWAS software according to He et al. (2017). In the MCSCP population, 277,581 SNPs were identified and further organized into 28,066 SNPLDBs. The SNP distribution on chromosomes of the MCSCP population was shown in **Supplementary Figure 1B**.

Genetic Richness, Similarity, and Differentiation of the MCSCP

The genetic diversity of the MCSCP was evaluated in two aspects: genetic richness (or the number of different alleles) and allele frequency dispersion on a locus. In the present study, the SNPLDB markers with their haplotypes were used to evaluate the genetic diversity, here the number of different haplotypes represents genetic richness while allele frequency dispersion was calculated (Carlson et al., 2005),

$$\pi = \frac{n}{n-1} \sum_{i=1}^q p_i (1 - p_i),$$

where n is the capacity of the population, p_i is the frequency of the i th allele/haplotype on a locus, q is the total allele number. The π of a SNPLDB was calculated based on the frequency of alleles in SNPLDBs, then the average of π of all SNPLDB in the population was used to represent the genetic dispersion of the population.

Wen et al. (2009) defined the SPA (specific present allele) of a population as an allele present in the population and SDA (specific deficient allele) as an allele present in all other populations but absent in the population. These two indicators were used to evaluate the specificity of the three subpopulations.

Genetic similarity (s_{ij})/distance (d_{ij}) was calculated for all pairs of cultivars: $d_{ij} = 1 - s_{ij}$, where s_{ij} is the coefficient of genetic similarity (CGS), calculated as $s_{ij} = \sum_{k=1}^m C_{ijk} / (2m)$, where C_{ijk} is the common allele number of the i th and j th cultivar at the k th SNPLDB marker, and m is the total number of SNPLDB markers. The CGS matrix was calculated based on all SNPLDBs using the RTM-GWAS software (He et al., 2017). The R software package (ECODist) was used for the Mantel test to analyze the correlation between COP matrix and CGS matrix of the MCSCP population (Goslee and Urban, 2007).

Principal component analysis was performed using RTM-GWAS (He et al., 2017) based on SNPLDB of the MCSCP. The first three principal components were used to analyze the relationship of population structure among three subpopulations of China.

The differentiation coefficient F_{ST} statistics between two subpopulations were calculated using the VCFtools software (Danecek et al., 2011) through the formula proposed by Weir and Cockerham (1984).

$$F_{ST} = \frac{\sum_u a_u}{\sum_u (a_u + b_u + c_u)},$$

where a_u , b_u , and c_u are the variances of frequencies for the u th allele between subpopulations, between individuals within subpopulations, and between gametes within individuals, respectively.

Genetic Clustering

Genetic clustering was performed using the genetic distance matrix of d_{ij} for both the 560 cultivars and the 11 major core-terminal ancestor-derived families using the neighbor-joining procedure of MEGA X (Kumar et al., 2018). The neighbor-joining tree of the 560 cultivars was generated from Figtree version 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

RESULTS

Agronomic Characteristics of the Modern Chinese Soybean Cultivars, Their Variability, and Differentiation Among Three Eco-Regions in China

Under the uniform environment testing at Dangtu and Jiaxiang, China, respectively, the MCSCP was performed to cover MG 000–MG IX (except MG X). All the six agronomic traits showed very significant variation with its DTF (days to flowering) ranging from 23.20 to 82.75 d, DTM (days to maturity) ranging from 63.25 to 144 d, and MSN (main stem node number) in 7.38–27.90 nodes, 100-seed weight in 7.17–40.26 g, seed protein content in 34.8–49.53% and seed oil content in 16.4–25.2% averaged over the two locations (**Table 1**). The results of the analysis of variance showed that there were significant differences among cultivars, environments, and interactions between cultivars and environments in the MCSCP (**Supplementary Table 3**).

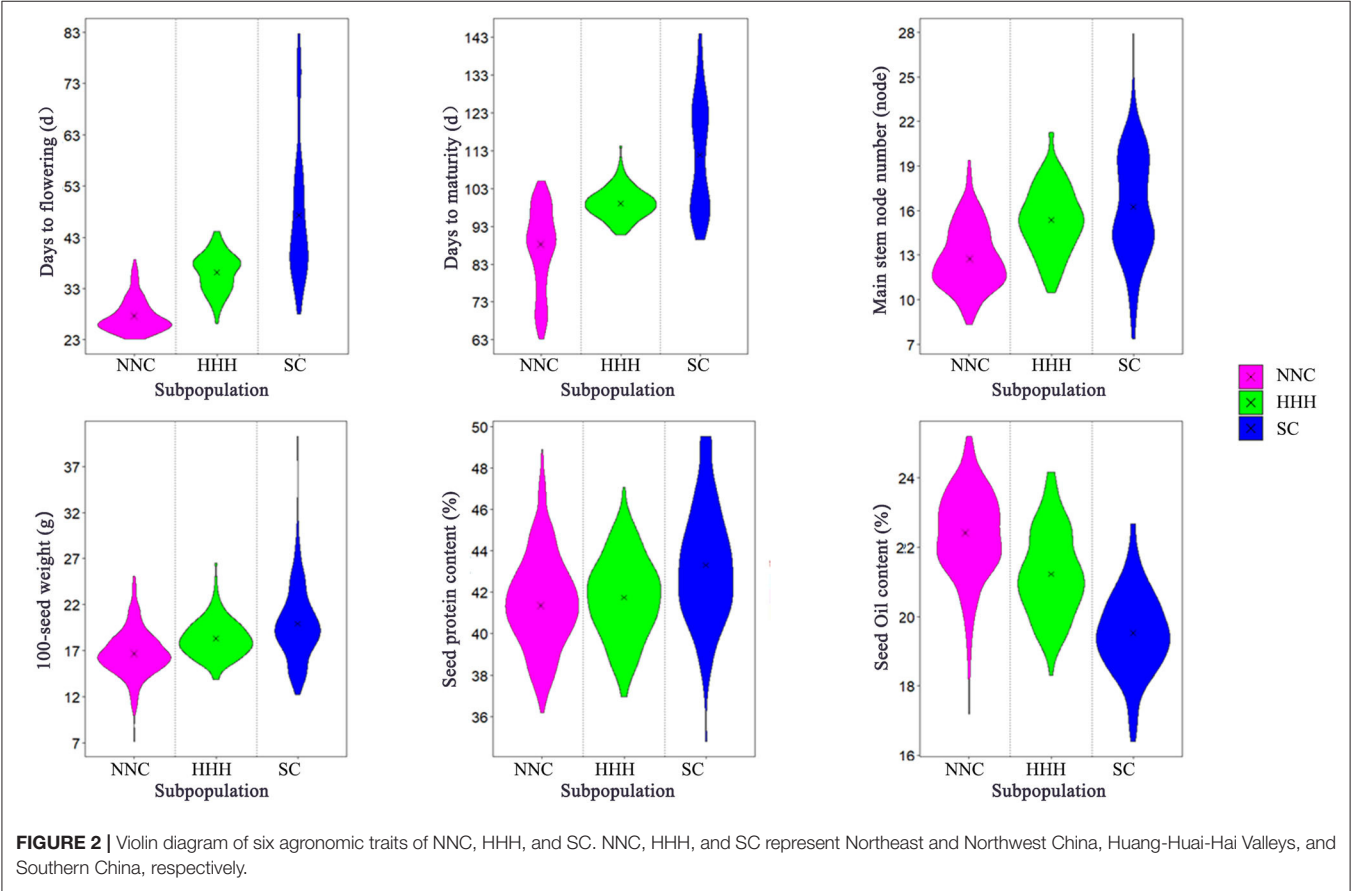
These agronomic traits varied among the three eco-regional subpopulations, especially those in SC under multiple cropping systems, its maturity groups covering MG I–IX, wider than the other two subpopulations (MG 000–IV in NNC and MG I–V in HHH). Its average DTF and DTM were 47.23 d (ranging from 28 to 82.75 d) and 111.68 d (ranging from 89.4 to 144 d), respectively, longer than the other two subpopulations (**Table 1**, **Figure 2**). Its average MSN and 100-seed weights was 16.31 nodes (ranging in 7.38–27.9 nodes) and 19.86 g (ranging in 12.25–40.26 g), more than the other two subpopulations. Its average protein content (43.31%, ranging from 34.8 to 49.53%) was significantly higher while the oil content (19.58%, ranging from 16.4 to 22.68%) was obviously lower than the other two subpopulations (**Table 1**, **Figure 2**). On the other hand, in the modern released cultivars in the HHH ecoregion, the maturity group varied from MG I to MG V, fitted in the double cropping system, while for those in the NNC region the MG varied in MG 000–MG IV, fitted in their full-season single cropping system. The different geographic locations caused the different cropping systems, in turn, caused the different maturity group constitutions in the three soybean ecoregions. For the agronomic traits, the growth-related traits, MG, DTF, DTM, and MSN, varied especially wide among the geographic regions due to the systematically changed photo-thermal conditions. While for the seed traits, 100-seed weight, protein content, and oil content varied with the selection pressure from farmers and breeders, in which, 100-seed weight varied wider than the other two traits because artificial selection pressure could act on seed size by sight in the long history.

TABLE 1 | Variability of agronomic traits within/among NNC, HHH, and SC of the MCSCP tested in Jiaxiang and Dangtu, China.

Ecoregion	No. of cultivars	Days to flowering			Days to maturity			Main stem node number		
		Mean (d)	Range (d)	GCV (%)	Mean (d)	Range (d)	GCV (%)	Mean (node)	Range (node)	GCV (%)
NNC	279	27.64c	23.20–38.60	11.09	87.62c	63.25–105.00	11.23	12.80c	8.32–19.36	15.18
HHH	155	36.15b	26.20–44.20	9.85	99.05b	90.80–114.25	2.15	15.36b	10.48–21.26	13.57
SC	126	47.23a	28.00–82.75	25.33	111.68a	89.40–144.00	12.05	16.31a	7.38–27.90	20.88
MCSCP	560	34.42	23.20–82.75	28.74	96.34	63.25–144.00	13.85	14.31	7.38–27.90	19.64

Ecoregion	Maturity group (MG)	100-seed weight			Seed protein content			Seed oil content		
		Mean (g)	Range (g)	GCV (%)	Mean (%)	Range (%)	GCV (%)	Mean (%)	Range (%)	GCV (%)
NNC	000–IV	16.73c	7.17–25.07	13.13	41.59b	36.20–48.88	4.66	22.38a	17.20–25.20	4.76
HHH	I–V	18.28b	13.88–24.56	5.47	41.59b	36.94–47.06	3.90	21.31b	18.85–24.16	4.74
SC	I–IX	19.86a	12.25–40.26	13.82	43.31a	34.80–49.53	5.36	19.58c	16.40–22.68	3.97
MCSCP	000–IX	17.83	7.17–40.26	13.38	41.97	34.80–49.53	6.79	21.45	16.40–25.20	6.32

Ecoregion: NNC, HHH, and SC represent Northeast and Northwest China, Huang-Huai-Hai Valleys, and Southern China, respectively. MCSCP means the modern Chinese soybean cultivar population released in 2006–2015.
MG, maturity group; GCV, genotypic coefficient of variation.
Different letters of a, b, and c in Column Mean indicate the means are significant at $p \leq 0.01$.



Germplasm Sources and Core-Terminal Ancestors of Modern Chinese Soybean Cultivars Based on Pedigree Analysis

The genetic background of the MCSCP was traced, starting from establishing the parental pedigree of an individual cultivar, from which, the immediate parents recognized, in turn, the terminal ancestors of a cultivar, the core-terminal ancestors on the whole population-level and ecoregion-level identified.

Nearly all cultivars (93.39%) of the MCSCP were developed through hybridization breeding (**Supplementary Table 4**). A total of 718 immediate parents were involved in MCSC development (**Figure 3D**), of which 74.09% were used only once. Only 17 direct parents were used more than six times (**Supplementary Table 5**). The major immediate parental types of the three subpopulations were cultivars and breeding lines, and the proportion of landrace and foreign variety as immediate parents in SC was significantly more than in the other two subpopulations (**Supplementary Table 5**). Compared to the proportion of immediate parents of cultivars released in 1923–2005 in the three regions, the proportion of immediate parents used only in southern cultivars was increased by 6.78%, while decreased by 4.09 and 2.83% in HHH and NNC, respectively, and the proportion of immediate parents shared by two or three regions changed slightly (**Figures 3A,D**).

The 560 cultivars were traced back to 604 terminal ancestors, of which 50 were shared by the three subpopulations and 70 were shared by two subpopulations (**Figure 3E**). Compared to those in the three regions released in 1923–2005, the proportion of the terminal ancestors used uniquely in their respective local ecoregions was about the same as in 1923–2005, and those shared by two or three regions increased (**Figures 3B,E**).

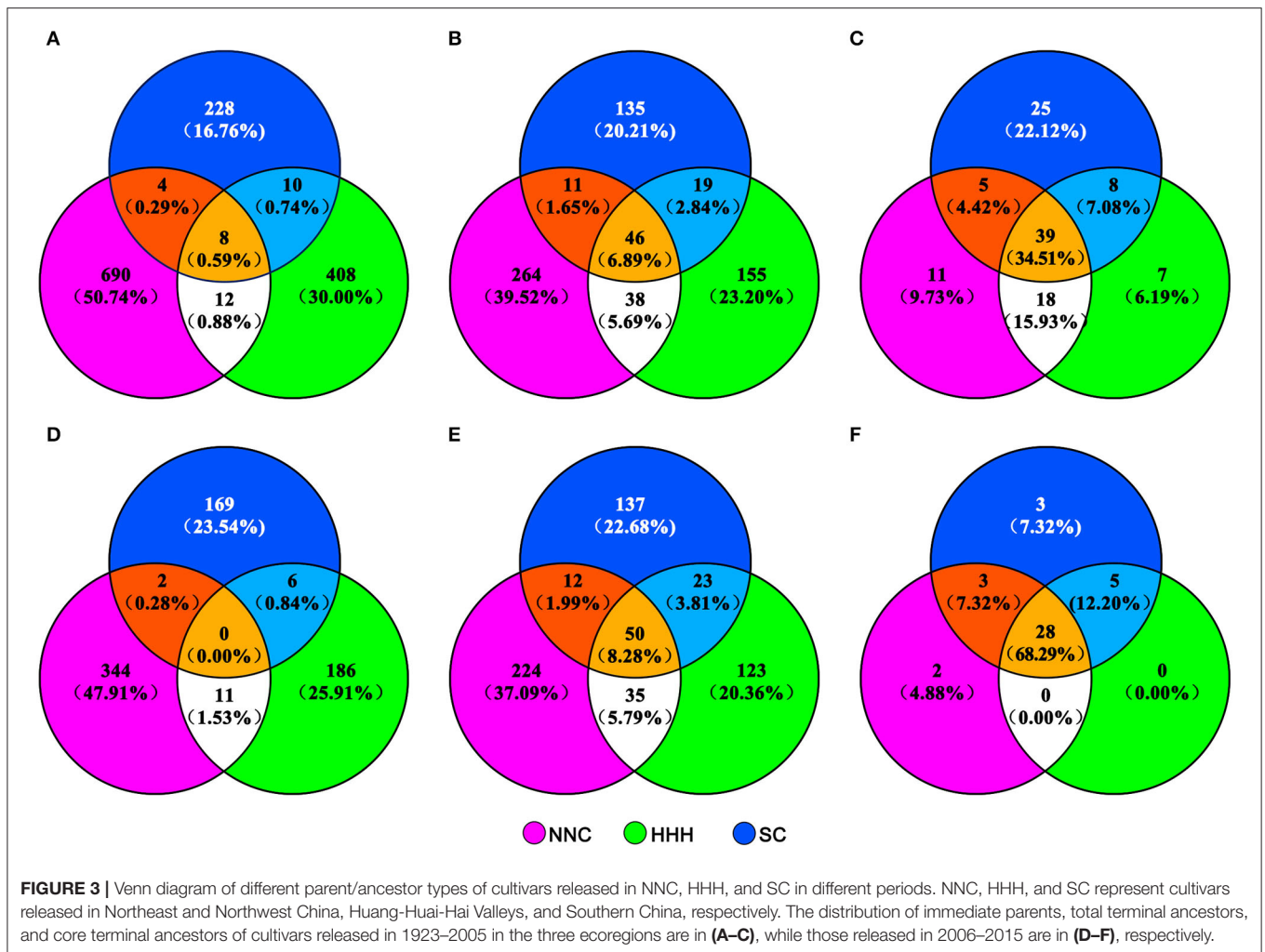
For the MCSCP, a total of 41 whole population-level core-terminal ancestors were identified from the 604 terminal ancestors, depending on their utilization in multiple breeding cycles, derived number of cultivars, nuclear contribution, and cytoplasmic contribution in the population as indicated above (**Table 2**). Their names with corresponding codes are shown in **Supplementary Table 6**. As an example, in Nannong 32 in **Figure 1**, there are two immediate parents and seven terminal ancestors, the nuclear contribution value of the seven terminal ancestors was added to 1 and the cytoplasmic contribution value was 1 from Fengxiansuidaohuang (CA-27). Among the seven terminal ancestors, three ones, Fengxiansuidaohuang (CA-27), 51-83 (CA-28), and Tongshantianedan (CA-16) are population-level core-terminal ancestors (**Figure 1**, **Table 2**). The 41 population-level core-terminal ancestors, accounting for 6.79% of the total terminal ancestors, contributed 44.47% to the total nuclear genetic contribution and 54.47% to the total cytoplasmic genetic contribution of the MCSCP (**Supplementary Table 7**). These population-level core-terminal ancestors comprise 14 ones that originated from NNC, 11 from HHH, nine from SC, and seven from abroad (**Table 2**, **Supplementary Table 6**). Compared to the population-level core-terminal ancestors of soybean cultivars released between 1923 and 2005 (Xiong et al., 2007), all 41 core ancestors are included in the 113 previous population-level core-terminal ancestors but varied in their

genetic contribution values. The genetic contribution changes of the core-terminal ancestors might be due to the shift of breeding objectives and the infiltration of new ones.

Among the 41 core-terminal ancestors, 33, 33, and 39 core-terminal ancestors are involved in NNC, HHH, and SC subpopulations, respectively, while 28 core terminal-ancestors are shared by all the three subpopulations (**Table 2**, **Figure 3F**). In the NNC ecoregion, about 1/4 of the nuclear genetic contribution came from the top five core terminal-ancestors (CA-01,–02,–03,–04, and–34), and more than half of cytoplasmic genetic contribution from the top five core terminal-ancestors (CA-01,–03,–05,–06, and–09), which are important terminal-ancestors in this region. In the HHH ecoregion, about 1/5 of nuclear genetic contribution from the top five core terminal-ancestors (CA-15,–16,–17,–18, and–36) and nearly half cytoplasmic genetic contribution from the top five core terminal-ancestors (CA-15,–17,–20,–21, and–36), which are important terminal-ancestors in this region. In SC ecoregion, about 1/10 of nuclear genetic contribution from the top five core terminal-ancestors (CA-15,–26,–27,–28, and–29) and nearly half cytoplasmic genetic contribution from the top five core terminal-ancestors (CA-15,–26,–27,–29, and–33), which are important terminal-ancestors in this region (**Table 2**). Compared to the proportion of whole population-level core terminal-ancestors of cultivars released in 1923–2005 in the three ecoregions, the proportion of core-terminal ancestors monopolized in SC, and the proportion of core-terminal ancestors shared by NNC and HHH decreased greatly, while those shared by NNC, HHH, and SC increased greatly (**Figures 3C,F**). Compared to the proportions of the immediate parents, terminal ancestors, and core-terminal ancestors in the three ecoregions from 1923 to 2005, those in NNC and HHH decreased, and those shared by HHH and SC increased to different degrees, indicating that in the breeding period of 2006–2015, the germplasm from HHH and NNC ecoregions contributed to the released cultivars in SC.

Genetic Richness and Differentiation of the MCSCP Based on Genomic Marker Analysis

Based on whole-genome sequencing, a total of 28,066 SNPLDBs with their 84,069 haplotypes were assembled and used in the genetic constitution analysis of the MCSCP with an SNPLDB-haplotype looked like a locus-allele. The total genetic richness (or the total number of alleles) of the MCSCP was 84,069 alleles. While the genetic richness of NNC, HHH, and SC subpopulations was 74,386 (88.48%), 72,894 (86.71%), and 72,843 (86.65%) alleles on the 28,066 loci, each locus with 2.65, 2.60, and 2.60 alleles, ranging in 2–19, 2–17, and 2–15 alleles per locus, respectively, indicating NNC with more genetic richness than the others (**Table 3**). Correspondingly, the allele-load per cultivar was 266.62 (74,386/279), 470.28 (72,894/155), and 578.12 (72,843/126) in NNC, HHH, and SC, respectively, implying more genetic difference among cultivars in latter ecoregions. On the other hand, the genetic dispersion among alleles on a locus (π) of NNC (0.106) was lower than that of HHH (0.129), and in turn



lower than that of SC (0.146), which means more allele frequency dispersion on the locus in HHH, and further in SC (Table 3).

To measure the genetic differentiation among subpopulations, the specific-present and specific-deficient alleles are substantial indicators. Here, a total of 7,537 SPAs and 16,988 SDAs were detected in the MCSCP, in which HHH had the smallest number of SPA (560) and SDA (4,198), many less than those in SC (2,374 and 6,061), and in turn less than those in NNC (4,603 and 6,729). The shared alleles between HHH and NNC (80.28%) and between HHH and SC (83.39%) were more than that between NNC and SC (76.3%), indicating genetically NNC was more distant from the other two subpopulations. It might be due to the single cropping system in NNC while double cropping in HHH and multiple cropping in SC (Table 3).

The genetic differentiation among ecoregions may be measured also using genetic similarity (s_{ij}) because the genetic distance is in fact $d_{ij} = 1 - s_{ij}$. Here we use average CGS to measure the population genetic differentiation, NNC was less distant to HHH (more similar as 87.32%) in comparison to SC (86.12%) and less distant to that between HHH and SC (85.66%). While in COP from pedigree analysis (larger COP means smaller genetic differentiation) and coefficient of

differentiation F_{ST} with the allele frequency included (larger F_{ST} means larger genetic differentiation), the results indicated that the genetic differentiation between HHH and NNC was lower (0.23 and 8.38%) than that between NNC and SC (0.12 and 10.45%). However, the differentiation between HHH and SC was the lowest (0.37 and 4.25%), which is somewhat different from the results in CGS that high differentiation in CGS but low differentiation in COP and F_{ST} (Table 3).

In summary, from the genetic constitution analysis, the NNC subpopulation was richer in allele number and SPA/SDA alleles, and genetically more distant from the other two subpopulations. The relationship among the three subpopulations was just as the principal component analysis showed that the cultivars of NNC (with a wide variation) connected with those of HHH and then extended to SC (Supplementary Figure 2).

Genetic Differentiation of the MCSCP in Terms of Genetic Clustering

The MCSCP differentiated among geographic regions, whether this kind of differentiation coincided with genetic differentiation was checked through the Neighbor-joining cluster analysis. The 560 cultivars were clustered into seven groups designated as A,

TABLE 2 | The core-terminal ancestors of the MCSCP as well as NNC, HHH and SC subpopulations.

Code	MCSCP			NNC			HHH			SC		
	NGC(%) ^a	CGC(%) ^b	No. C	NGC (%)	CGC (%)	No. C	NGC (%)	CGC (%)	No. C	NGC (%)	CGC (%)	No. C
CA-01	3.04 (3.52)	4.46 (6.38)	195	6.02	8.96	190	0.04		2	0.15		3
CA-02	2.56 (5.05)	0.00 (1.31)	296	4.78		233	0.48		47	0.19		16
CA-03	2.41 (4.46)	10.18 (10.38)	272	4.49	20.43	225	0.39		38	0.26		9
CA-04	2.05 (1.83)	0.00 (0.08)	160	4.06		157	0.03		1	0.07		2
CA-05	2.04 (2.78)	3.57 (3.54)	220	3.15	7.17	150	1.16		50	0.66		20
CA-06	1.66 (1.55)	5.18 (2.00)	109	2.84	9.68	87	0.42	0.65	15	0.57	0.79	7
CA-07	1.31 (1.23)		190	2.12		126	0.45		44	0.57		20
CA-08	1.17 (2.22)	0.00 (0.08)	191	1.96		140	0.44		36	0.31		15
CA-09	1.04 (0.89)	5.00 (4.46)	101	2.02	10.04	98	0.04		1	0.11		2
CA-10	0.85 (1.08)	0.18 (0.54)	82	1.62		81				0.2	0.79	1
CA-11	0.80 (0.78)		75	1.56		73				0.11		2
CA-12	0.71 (0.51)	0.00 (0.31)	96	1.43		96						
CA-13	0.65 (0.47)	0.00 (0.08)	51	1.31		51						
CA-14	0.36 (0.62)	0.00 (0.46)	118	0.38		73	0.39		31	0.27		14
CA-15	2.41 (2.87)	4.64 (3.92)	144	0.14	0.72	5	6.94	12.26	110	1.9	3.97	29
CA-16	1.12 (1.38)	0.18 (0.92)	127	0.02		2	3.43	0.65	101	0.72		24
CA-17	1.06 (0.57)	2.50 (1.23)	44				3.63	8.39	40	0.24	0.79	4
CA-18	0.98 (0.52)		64	0.09		1	3.26		59	0.15		4
CA-19	0.89 (1.63)	1.25 (2.15)	96	0.1		9	2.72	4.52	74	0.38		13
CA-20	0.82 (1.48)	1.61 (2.92)	113	0.17	0.36	13	2.22	5.16	83	0.53		17
CA-21	0.77 (0.43)	3.04 (2.00)	70	0.05		2	2.56	10.97	62	0.19		6
CA-22	0.75 (0.49)	0.00 (0.08)	75	0.05		2	2.46		67	0.22		6
CA-23	0.52 (0.33)	0.71 (0.92)	63	0.04		1	1.75	1.94	58	0.09	0.79	4
CA-24	0.39 (0.64)		107	0.09		13	1.06		81	0.23		13
CA-25	0.39 (0.77)	0.00 (0.23)	107	0.09		13	1.06		81	0.23		13
CA-26	1.36 (1.48)	3.39 (3.08)	32							6.05	15.08	32
CA-27	0.80 (0.98)	1.61 (3.46)	46				1.14	1.29	33	2.15	5.56	13
CA-28	0.52 (0.77)	0.71 (3.15)	43				1.16	1.94	36	0.91	0.79	7
CA-29	0.49 (0.72)	1.61 (2.62)	19	0.04	0.36	1				2.1	6.35	18
CA-30	0.31 (0.32)	0.00 (2.23)	28				0.96		23	0.19		5
CA-31	0.27 (0.10)	0.89 (1.31)	32				0.98	3.23	31	0.01		1
CA-32	0.16 (0.14)	0.00 (0.46)	8							0.69		8
CA-33	0.03 (0.22)	0.36 (0.62)	3							0.15	1.59	3
CA-34	3.16 (3.13)	0.36 (0.38)	186	6.22	0.72	181	0.2		4	0.05		1
CA-35	1.66 (1.80)	0.00 (0.46)	220	2.37		135	1.15		58	0.74		27
CA-36	1.19 (1.21)	3.04 (1.69)	135	0.45	0.72	44	2.82	9.68	77	0.83		14
CA-37	1.15 (1.14)		156	0.47		48	2.58		79	0.9		29
CA-38	0.79 (0.61)		184	1.08		124	0.76		53	0.21		7
CA-39	0.72 (0.94)	0.00 (0.08)	179	1.1		126	0.37		32	0.3		21
CA-40	0.56 (0.71)		127	0.01		2	1.72		101	0.36		24
CA-41	0.56 (0.71)		127	0.01		2	1.72		101	0.36		24
Total	44.47 (53.08)	54.47 (63.53)		50.33	59.16		50.49	60.68		24.35	36.50	

Code: the codes of core-terminal ancestors, and their corresponding names were shown in **Supplementary Table 6**.

MCSCP is the abbreviation for the modern Chinese soybean cultivar population. NNC, HHH, and SC represent cultivars released in Northeast and Northwest China, Huang-Huai-Hai Valleys, and Southern China, respectively.

NGC: the percentage of nuclear genetic contribution relative to the total released cultivars in the population; CGC: the percentage of cytoplasmic genetic contribution relative to the total released cultivars in the population; No. C: number of derived cultivars.

^a and ^b: in Column NGC and CGC, those in parentheses are the percentage of nuclear genetic contribution and cytoplasmic genetic contribution in the population (1300 released cultivars) released in 1923–2005, respectively.

TABLE 3 | Genetic richness, dispersion, similarity, and differentiation of NNC, HHH, and SC subpopulations in the MCSCP.

Subpopulation/ Population	No. of cultivars	Allele No.			π		SPA	SDA	Shared allele (%)		
		Total	Per locus	Max/Locus	Mean	Max			NNC	HHH	SC
NNC	279	74,386	2.65	19	0.106	0.889	4,603	6,729			
HHH	155	72,894	2.60	17	0.129	0.876	560	4,198	80.28		
SC	126	72,843	2.60	15	0.146	0.900	2,374	6,061	76.30	83.39	
MCSCP	560	84,069	3.00	21	0.128	0.909	7,537	16,988			

Subpopulation/ Population	COP ($\times 100$)			CGS ($\times 100$)			F _{ST} ($\times 100$)		
	NNC	HHH	SC	NNC	HHH	SC	NNC	HHH	SC
NNC	2.74			89.35					
HHH	0.23	2.11		87.32	87.08		8.38		
SC	0.12	0.37	1.15	86.12	85.66	85.35	10.45	4.25	

Subpopulation/Population: MCSCP represents the modern Chinese soybean cultivar population released in 2006–2015. NNC, HHH, and SC represent cultivars released in Northeast and Northwest China, Huang-Huai-Hai Valleys, and Southern China, respectively.

SPA, specific-present allele; SDA, specific-deficient allele.

Shared allele: the percentage of shared alleles to all alleles in the two population.

COP: coefficient of parentage within and between NNC, HHH, and SC subpopulations, the actual values were multiplied by 100 in the table.

CGS: coefficient of genetic similarity within and between NNC, HHH, and SC subpopulations, the actual values were multiplied by 100 in the table.

F_{ST}: fixation index between NNC, HHH, and SC subpopulations, the actual values were multiplied by 100 in the table.

B, to G (**Figure 4A**, **Table 4**). The NNC cultivars were clustered in A, B, C, and D groups; while the HHH cultivars were in B to F groups and SC cultivars were in B to G groups. The three geographic subpopulations had a different distribution among the genetic clusters which means the subpopulations had their major-specific cluster(s) but shared some clusters. All 150 cultivars in Group A were from NNC, while most of the cultivars in B (108) were also from NNC with the others from HHH and SC. In C, D, and E groups, the major ones were from HHH, while some in C, and D form NNC and the others from SC. In the F group, the major ones were from SC while the others were from HHH; but all the 62 cultivars in the G group were purely from SC. From the above, the clustering results indicated the geographic grouping was not completely but partially consistent with the genetic clustering. Therefore, in each subpopulation, there exist different genetic sources. This understanding also coincided with the geographic distribution of the ancestor-derived Cultivar families in the subsequent text.

Genetic Differentiation of Modern Chinese Soybean Cultivars in Terms of Ancestor-Derived Cultivar Families

According to the progeny-pedigrees of the population-level core-terminal ancestors, those with the largest numbers of derived cultivars and high nuclear and cytoplasmic genetic contributions were recognized as major core-terminal ancestors, which along with its derived cultivars, were called a core-terminal ancestor-derived family. In this way, 11 major core-terminal ancestor-derived families were identified in the MCSCP. Their core-terminal ancestors were traced to CA-01 (Baimei), CA-02 (Jinyuan), CA-03 (Silihuang), and CA-05 (Duludou) from

NNC, CA-15 (Binhaidabaihua), CA-16 (Tongshantianedan), CA-19 (Shouzhangdifangzhong) and CA-20 (Jimoyoudou) from HHH, CA-26 (Vegetable soybean population in Wuhan), CA-27 (Fengxiansuidaohuang) and CA-28 (51-83) from SC. These 11 major core-terminal ancestors out of the 604 terminal ancestors provided $100.65/560 = 17.97\%$ nuclear genetic contribution and $177/560 = 31.6\%$ cytoplasmic contribution to the MCSCP (**Table 5**).

From the 11 major core-terminal ancestors, 463 cultivars were derived, and each cultivar involved one or more major core-terminal ancestors. Therefore, a total of 1,584 cultivar • times involved with the 11 major core-terminal ancestors. Among the 1,584 cultivar • times, 983 ones were from the four NNC major core-terminal ancestors, in which 798 cultivar • times for NNC cultivars, 137 cultivar • times for HHH cultivars and 48 cultivar • times for SC cultivars. While among the 1,584 cultivar • times, 480 cultivar • times were from the four HHH major core-terminal ancestors, in which 368 cultivar • times for HHH cultivars, 29 cultivar • times for NNC cultivars, and 83 cultivar • times for SC cultivars. Furthermore, among the 1,584 cultivar • times, 121 cultivar • times were from the three SC major core-terminal ancestors, in which 52 cultivar • times for SC cultivars, 69 cultivar • times for HHH cultivars, and none for NNC cultivars. Thus, there showed germplasm exchange among the three ecoregions regarding the 11 major core-terminal ancestor-derived families, the NNC germplasm and HHH germplasm penetrated to the other two regions, but no SC germplasm to NNC, especially CA-26 even not to HHH. In the 560 MCSCs, $56.24/560 = 10.04\%$ nuclear and $102/560 = 18.21\%$ cytoplasmic germplasm were from the four NNC major core-terminal ancestors, $29.38/560 = 5.25\%$ nuclear and $43/560 = 7.68\%$ cytoplasmic germplasm were from the four HHH major

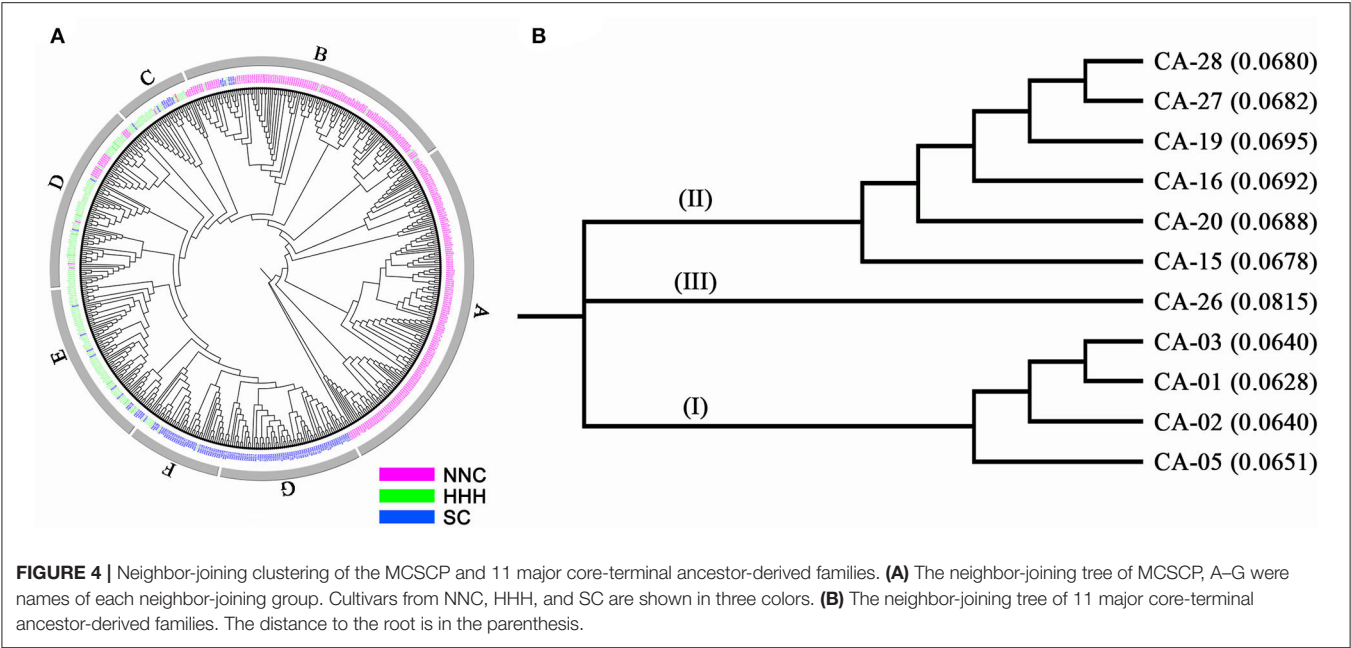


TABLE 4 | The distribution of NNC, HHH, and SC cultivars in seven genetic clusters.

Subpopulation	Neighbor-joining cluster group							Total
	A	B	C	D	E	F	G	
NNC	150	108	4	17				279
HHH		5	20	64	60	6		155
SC		7	8	2	12	35	62	126
Total	150	120	32	83	72	41	62	560

NNC, HHH, and SC represent cultivars released in Northeast and Northwest China, Huang-Huai-Hai Valleys, and Southern China, respectively.

TABLE 5 | Frequency distribution among ecoregions and genetic parameters of the 11 major core-terminal ancestor-derived families.

Code	Distribution of derived cultivars				NGC	CGC	Allele No.				π	
	NNC	HHH	SC	Total			Total	Per-locus	Per-cultivar	Max/Locus	Mean	Max
CA-01	190	2	3	195	17.05	25	68,903	2.46	353.35	16	0.103	0.880
CA-02	233	47	16	296	14.32		81,909	2.92	276.72	21	0.114	0.906
CA-03	225	38	9	272	13.45	57	80,328	2.86	295.32	20	0.112	0.899
CA-05	150	50	20	220	11.42	20	81,790	2.91	371.77	21	0.120	0.918
CA-15	5	110	29	144	13.52	26	74,447	2.65	516.99	17	0.134	0.891
CA-16	2	101	24	127	6.29	1	72,803	2.59	573.25	16	0.134	0.882
CA-19	9	74	13	96	4.98	7	70,622	2.52	735.65	16	0.133	0.894
CA-20	13	83	17	113	4.59	9	73,412	2.62	649.66	18	0.135	0.918
CA-26			32	32	7.63	19	55,245	1.97	1726.41	11	0.140	0.881
CA-27		33	13	46	4.47	9	60,367	2.15	1312.33	13	0.133	0.901
CA-28		36	7	43	2.93	4	56,760	2.02	1320.00	12	0.129	0.904
Total	253(827)	137(574)	73(183)	463(1,584)	(100.65)	177	83,966	2.99	181.35	21	0.125	0.906

Code: the code of the core terminal ancestors, and the corresponding name of the core terminal ancestors were shown in **Supplementary Table 6**. The derived cultivars may duplicate among the core-terminal ancestor-derived families, so the total cultivars were 463 while the total cultivar • time was 1,584 (in parentheses) and the same is for each ecoregion and for NGC (not for CGC). Per-cultivar means per-cultivar allele load calculated from the total allele number divided by the number of derived cultivars, which is a measure of the allele richness in a core-terminal ancestor-derived family.

core-terminal ancestors, and only $15.03/560 = 2.68\%$ nuclear and $32/560 = 5.71\%$ cytoplasmic germplasm were from the three SC major core-terminal ancestors. It seems that both nuclear and cytoplasmic germplasm of the core-terminal ancestor-derived families might dominate in the MCSCP (Table 5).

As for the allele constitution, the total number of alleles in the 11 major ancestor-derived families ranged from 55,245 to 81,909, and the richness of NNC major ancestor-derived families (in an average of 78232.5) was more than those of HHH ones (in an average of 72,821), and in turn more than those of SC (in an average of 57457.3). The same tendency was for per-locus allele number (2.46–2.92, 2.52–2.65, and 1.97–2.15 for NNC, HHH, and SC major ancestor-derived families, respectively) and allele frequency dispersion (0.103–0.12, 0.133–0.135, and 0.129–0.14 for NNC, HHH, and SC major ancestor-derived families, respectively) but in the opposite situation for per-cultivar allele load (276.72–371.77, 516.99–735.65, and 1312.33–1726.41 for NNC, HHH, and SC major ancestor-derived families, respectively). That means the NNC major ancestor-derived families are richer in their allele sources and per-locus allele differentiation but are poorer in their per-cultivar allele load than that of the other two major ancestor-derived family groups (Table 5).

The results of neighbor-joining clustering showed that the 11 major ancestor-derived families were classified into three groups (Figure 4B). Group I is the most advanced cluster with the shortest distances to the root, which includes derived-families from CA-05, CA-02, CA-01, and CA-03, all the four major core-terminal ancestors were from NNC. Group II includes derived families from CA-15, CA-20, CA-16, and CA-19 (from HHH) CA-27, and CA-28 (from SC). This group included major core-terminal ancestors from both HHH and SC, indicating the relatively close relationship between the two sets of the ancestor-derived families. Group III includes the CA-26 derived family, which was the most primitive one with the longest distance to the root among the 11 families. CA-26 is a vegetable soybean population in SC.

Features of COP and CGS in Evaluating the Genetic Similarity/Diversity Among the Released Cultivars

In studying the genetic relationship among released cultivars, COP was used at first based on parental pedigrees, while after the molecular marker technique developed, CGS was used extensively. COP represents the probability of the two cultivars carrying the same allele inherited from the same parent, i.e., identity by descent (IBD) (Hallauer and Miranda, 1981). In contrast, CGS represents the probability of the two cultivars carrying the same allele, regardless of the same parent, i.e., identity by state (IBS) (Hallauer and Miranda, 1981). A larger COP or CGS indicates a closer genetic relationship between the two cultivars. For convenience, all the calculated values of COP and CGS in the present study were multiplied by 100. The COP of the MCSCP was averaged to 1.04, not closely genetic-related among cultivars in the MCSCP, varying from 0 to 100, with a large variation from non-genetic relationship to full genetic relatedness (Table 6). Among the subpopulations, the COP values of NNC,

HHH, and SC were averaged at 2.74, 2.11, and 1.15 respectively, while in NNC, it ranged from 0 to 70 and from 0 to 100 in HHH and SC. The COP of the most cultivar pairs concentrated in the groups of 0 and 0–10 (56.14 + 42.11% in the MCSCP, 20.90 + 74.01% in NNC, 29.37 + 67.52% in HHH, and 82.18 + 14.69% in SC), indicating NNC and HHH with more relatedness among cultivars than SC or more distant among cultivars in SC. In Table 6, the CGS of the MCSCP was averaged 87.23 in the MCSCs, varying from 74.62 to 98.73. Among the subpopulations, the CGS values were 89.35, 87.08, and 85.35 in NNC, HHH, and SC, ranging from 76.73 to 96.59, 81.02 to 98.73, and 79.48 to 96.18, respectively. In the MCSCP, the CGS of the most cultivar pairs concentrated in the groups of 80–90 and 90–100 (82.49 + 15.78% in MCSCP, 52.41 + 47.10% in NNC, 92.75 + 7.25% in HHH, and 98.41 + 1.35% in SC), indicating also NNC with more relatedness among cultivars than the latter two or more distant among cultivars in HHH and SC. Both COP and CGS confirmed that differentiation of genetic relatedness also happened among subpopulations due to geographic and agronomic differences. The shortened growth season caused a mono-/single-cropping system in NNC, an extended growth season caused a double-cropping system in HHH, and a long growth season caused multiple cropping systems in SC. Altogether happened the differentiation in requirements to cultivars and their parentage materials, therefore the genetic relatedness among cultivars.

The correlation between the COP and CGS matrices was $r = 0.229$, significant under the Mantel test in the MCSCP. Supplementary Table 8 showed the two-way table of COP and CGS frequency distribution. In the MCSCP, the most cultivar pairs concentrated in the four cells of COP groups of 0–0 and 0–2.5 by CGS groups of 84–89 and 89–94. The data indicated that the COP and CGS were correlated, but not closely, the two genetic relationship indicators have their own features. COP is characterized in that it is based on exact parental pedigrees, otherwise not possible to be calculated, especially, since the unknown relatedness between ancestors is not easy to be recognized (0 or 1 has to be assumed). In addition, it represents a theoretical genetic relationship useful in inheritance analysis, but not an exact one due to the fluctuation in segregation departing from the assumption of the equal genetic contribution of the parents to progenies. CGS is characterized that it is based on experimental data with full information from each cultivar/individual, especially, providing exact calculations coinciding with the real case, and that no assuming value (0 or 1) is required and easy in the algorithm, such as in GWAS procedures. Thus, in experimental genetics, CGS is more pragmatic than COP in studying the genetic relationship among/between cultivars.

DISCUSSION

Pedigree Integrated With Genomic-Marker (SNPLDB) Analysis Advanced the Germplasm Study

In the study of released cultivars as germplasm sources, pedigree analysis has been a major approach. For a cultivar, there are pedigrees of its parental ancestors and pedigrees of its

TABLE 6 | Frequency distribution of COP and CGS in NNC, HHH, and SC of the MCSCP.

Population	COP set											Mean	
	0	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100		Total
NNC	8,104 (20.90)	28,702 (74.01)	1,365 (3.52)	442 (1.14)	113 (0.29)	37 (0.10)	17 (0.04)	1 (0.00)				38,781 (100)	2.74
HHH	3,505 (29.37)	8,059 (67.52)	176 (1.47)	150 (1.26)	16 (0.13)	16 (0.13)	10 (0.08)			1 (0.01)	2 (0.02)	11,935 (100)	2.11
SC	6,472 (82.18)	1,157 (14.69)	96 (1.22)	97 (1.23)	17 (0.22)	26 (0.33)		3 (0.04)	1 (0.01)	3 (0.04)	3 (0.04)	7,875 (100)	1.15
MCSCP	87,873 (56.14)	65,903 (42.11)	1,731 (1.11)	734 (0.47)	150 (0.10)	88 (0.06)	27 (0.02)	4 (0.00)	1 (0.00)	4 (0.00)	5 (0.00)	15,6520 (100)	1.04
Population	CGS set											Mean	
	0	0-10	10-20	20-30	30-40	40-50	50-60	60-70	70-80	80-90	90-100		Total
NNC									189 (0.49)	20,326 (52.41)	18,266 (47.10)	38,781 (100)	89.35
HHH										11,070 (92.75)	865 (7.25)	11,935 (100)	87.08
SC									19 (0.24)	7,750 (98.41)	106 (1.35)	7,875 (100)	85.35
MCSCP									2,703 (1.73)	129,116 (82.49)	24,701 (15.78)	156,520 (100)	87.23

COR, coefficient of parentage; CGS, coefficient of genetic similarity. In the "Population" column, NNC, HHH, and SC represent cultivars released in Northeast and Northwest China, Huang-Huai-Hai Valleys, and Southern China, respectively. MCSCP represents the modern Chinese soybean cultivar population. The number of cultivar pairs is outside of the parentheses, while the percentage of frequency is in parentheses. In the "Mean" column, the number was the mean value of COP or CGS of the population, which was multiplied by 100.

progenies, here we called parental pedigree and progeny pedigree, respectively. In the present study, in tracing the released cultivars' parental ancestors, the parental pedigree analysis was utilized while in tracing core-terminal ancestor-derived families, the progeny pedigree analysis was applied.

In soybean, parental pedigree analysis was the major procedure in studying the genetic basis of released cultivars before the molecular marker and genomic analysis. Delannay et al. (1983) and Gizlice et al. (1994) analyzed the pedigrees of North American cultivars released in different periods and traced their ancestors. Cheng and Hadley (1986) and Bernard (1988) analyzed the sources, pedigrees, and ancestors of cultivars released in the US and Canada in different periods. Hiromoto (1986), Zhou et al. (2000), Bhardwaj et al. (2002), and Wysmierski and Vello (2013) analyzed the pedigree of Brazilian, Japanese, Indian, and Brazilian soybean cultivars, then summarized the main ancestors or core ancestors in a certain period, respectively. These were of great significance to germplasm exchange among different regions in broadening the genetic basis of cultivars. In the present study, 604 terminal ancestors were summarized from the 560 cultivars, and 41 core-terminal ancestors were extracted from them as the representatives of important terminal ancestors in the Chinese soybean germplasm population, and then the germplasm sources of the MCSCP were clarified from the perspectives of the parental pedigrees. In comparison to Xiong et al. (2007), the 41 presently nominated population-level core-terminal ancestors in 2006–2015 were included in those during 1923–2005, but their frequencies changed. That means the germplasm exchange among ecoregions during the recent period (2006–2015) did not shack the basic genetic background established in 1923–2005. In other words, the MCSCs have inherited the superior genes-alleles from those released in about a century, while the gene exchange through introducing new parental materials may have caused some changes due to the differences in inter-regional breeding activities but not very much (Figure 3). In the MCSCP, the core-terminal ancestors shared between HHH and SC increased by 5.12%, while that between HHH and NNC decreased by 15.93%. It indicates that the germplasm infiltration from HHH to SC is more common than from HHH to NNC, therefore, the germplasm exchange among subpopulations has been being appealed during recent years. On the other hand, the progeny pedigree analysis was based in fact on parental analysis of the derived cultivars, i.e., the 41 core-terminal ancestor and 11 major core-terminal ancestor-derived families were based on the parental pedigree analysis of 560 and 463 cultivars, respectively. From the analysis, the most important germplasm sources in the MCSCP were summarized.

Genetic markers helped to characterize the genetic constitution, the genetic difference among cultivars, and genetic differentiation among the subpopulations using genetic richness, SPA, SDA, allele frequency dispersion, and genetic similarity/diversity as indicated in the results. Using the marker-assisted analysis, the multiple germplasm sources of the geographic subpopulations were found through clustering analysis. Especially, the genomic markers SNPLDBs with multiple alleles were used which can cover the whole genome and provide relatively thorough genetic information. Therefore,

the pedigree analysis integrated with genomic-maker SNPLDB analysis advanced the present study on released cultivars as growing germplasm sources.

Geographic and Breeding-Effort Factors Involved in the Differentiation of the MCSCP and Implications for Breeding for Soybeans

In the results section, the differentiation of the MCSCP was analyzed from three aspects, i.e., geographic differentiation, genetic structure differentiation, and genetic clustering differentiation due to ancestor changes, which involved geographic factor and breeding-effort/parentage factor. At the early stage of the population differentiation, the geographic factors determined the differentiation direction for getting soybeans adaptive to local environments, including geographic photo-thermal and sowing seasonal conditions (cropping system). Later on, the exchange of parental materials caused the outside germplasm introduced into the local region. As indicated above, although the soybean production history was very long in China, the scientific breeding history was only about one century, while the germplasm exchange history was now at its early stage.

In comparison, the American soybean cultivars, which were developed based on mainly the introductions of Northeast China germplasm during a recent couple of centuries in their scientific breeding plans, but the American cultivars have their yield potential exceeding their Chinese ancestors. Liu et al. (2020) reported that in the global clustering of the 13 geographic soybean subpopulations, those of the north and south Americas were clustered with those of Northeast China in the same group, while the latter was not with the Asian group. Now the American soybeans have covered all 13 MGs, almost the same as the Asian group. The Asian group took a long history to get soybean cultivars developed under a relatively self-sufficient manner in using germplasm while the American group took a very short period to get soybean cultivars developed under a broad exchanging manner in using germplasm. This is an obvious example of the global differentiation of soybeans due to the breeding-effort/parentage factor in addition to the geographic factor.

Gizlice et al. (1994) found that 28 ancestors and seven first progenies contributed 95.97% of the genes found in public cultivars released between 1947 and 1988 in North America. In these 35 cultivars, four ancestors (Mandarin, Richland, A.K., and Dunfield) and one derived-progeny (Lincoln) contributed 46.73% of the genes. The five accessions also existed in the 41 core-terminal ancestors of 2006–2015 and in the 113 core-terminal ancestors of 1923–2005 in the Chinese soybean population (Xiong et al., 2007), providing 5.51 and 5.7% germplasm contribution, respectively. This indicates that the germplasm from outside of China should be introduced back and utilized in soybean improvement in the three eco-regions because the outside germplasm has been improved by foreign breeders even though its source was initially based on the Chinese introductions.

In essence, breeding is a process of introducing and accumulating superior genes/alleles of target traits and eliminating inefficient genes/alleles from different parents. Wild soybeans and landraces generally contain excellent or special trait genes, but their comprehensive genetic background is poor and unsuitable to be directly used in production, so they are only suitable as donor parents. Released cultivars are an important and the most widely used type of soybean germplasm resource, especially the recently released cultivars, which have been improved for several breeding cycles and can meet the production demands at the present stage. Therefore, released cultivars are important receptor materials to receive complimentary superior genes/alleles from donors for further improvement in the future.

The above results showed that the germplasm sources of the three regions are relatively independent, except for some small amount of exchange and utilization happened. For further soybean improvement in China, the germplasm exchange among ecoregions should be enhanced and at the same time, the introduction of excellent germplasm resources from abroad should be emphasized also for adding new alleles or allele-recombinants to generate new genetic types, broadening the genetic base of the local population and enriching the genetic diversity of future cultivars.

CONCLUSIONS

The modern released cultivar population is a genetic deposition of historical accumulation of breeding effort or superior genes-alleles. The MCSCP, composed of 560 released cultivars (279 in NNC, 155 in HHH, and 126 in SC), covered a wide range of geo-seasonal ecotypes in terms of maturity groups (MG 000–MG IX). Its germplasm sources were traced to 718 immediate parents, 604 terminal ancestors, and 41 core-terminal ancestors through parental-pedigree analysis of the released cultivars. Among the three eco-subpopulations, the germplasm sources were different from each other with a part shared by two or three eco-subpopulations (2.65% immediate parents, 19.87% terminal-ancestors), while 87.81% core-terminal ancestors shared by two or three eco-subpopulations. The genomic marker analysis further indicated the MCSCP differentiation among the three subpopulations, NNC with more genetic richness (74,386 alleles vs. 72,894 and 72,843 alleles), more SPA, and SDA than the others (4,603 and 6,729 vs. 560 and 4,198, and 2,374 and 6,061 alleles, respectively). In evaluating genetic relationship/differentiation, CGS is more pragmatic than COP. The CGS results confirmed NNC was less distant from HHH (CGS 87.32%) in comparison to SC (CGS 86.12%) and less distant than to that between HHH and SC (CGS 85.66%). The genetic clustering of the MCSCP further supported the genetic differentiation among the three subpopulations but was partially caused by the interchange of genetic sources. Combining the parental-pedigree with genomic marker analysis, the major core-terminal ancestor-derived families are also characterized by eco-regional differentiation. All the results showed that the germplasm sources of the three eco-subpopulations are relatively self-sufficient, germplasm exchange

only at an early stage, for further improvement, domestic and international germplasm exchange is seriously required.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**; the raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

CLi, JZ, and JG conceived the study and drafted, revised, and finalized the manuscript. CLi, JZ, CLiu, XZ, and JC collected the cultivars with their pedigrees. CLi, CLiu, XZ, JC, and WW performed the experiment. CLi, WW, YP, JH, and FL performed data analysis. All authors approved the submitted manuscript.

REFERENCES

- Andolfatto, P., Davison, D., Erezylmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., et al. (2011). Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res.* 21, 610–617. doi: 10.1101/gr.115402.110
- Bernard, R. L. (1988). *Origins and Pedigrees of Public Soybean Varieties in the United States and Canada*. Washington, DC: U.S. Department of Agriculture. Technical Bulletin 1746, 01–68.
- Bhardwaj, C., Satyavathi, C. T., Tiwari, S., and Karmakar, P. (2002). Genetic base of soybean (*Glycine max*) varieties released in India as revealed by coefficient of parentage. *Indian J. Agric. Sci.* 72, 467–469.
- Browning, B. L., Zhou, Y., and Browning, S. R. (2018). A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* 103, 338–348. doi: 10.1016/j.ajhg.2018.07.015
- Carlson, C. S., Thomas, D. J., Eberle, M. A., Swanson, J. E., Livingston, R. J., Rieder, M. J., et al. (2005). Genomic regions exhibiting positive selection identified from dense genotype data. *Genome Res.* 15, 1553–1565. doi: 10.1101/gr.4326505
- Cheng, S., and Hadley, H. (1986). *Cytoplasm Sources of Soybean Cultivars Grown in the United States and Canada*. Soybean genetics newsletter. Washington, DC: United States Agricultural Research Service (USA).
- Cox, T., Kiang, Y., Gorman, M., and Rodgers, D. (1985). Relationship between coefficient of parentage and genetic similarity indices in the soybean 1. *Crop Sci.* 25, 529–532. doi: 10.2135/cropsci1985.0011183x002500030023x
- Cui, Z., Carter, T. E., and Burton, J. W. (2000). Genetic base of 651 chinese soybean cultivars released during 1923 to 1995. *Crop Sci.* 40, 1470–1481. doi: 10.2135/cropsci2000.4051470x
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Delannay, X., Rodgers, D., and Palmer, R. (1983). Relative genetic contributions among ancestral lines to North American soybean cultivars¹. *Crop Sci.* 23, 944–949. doi: 10.2135/cropsci1983.0011183x002300050031x
- Fehr, W., and Caviness, C. (1977). *Stages of Soybean Development*. Special Report 80. Ames, IA: Iowa Agricultural Experiment Station, Iowa Cooperative External Series, Iowa State University.
- Fu, M., Wang, Y., Ren, H., Du, W., Yang, X., Wang, D., et al. (2020). Exploring the QTL-allele constitution of main stem node number and its differentiation among maturity groups in a Northeast China soybean population. *Crop Sci.* 60, 1223–1238. doi: 10.1002/csc2.20024
- Fu, M. M., Wang, Y. P., Ren, H. X., Wang, D. L., Bao, R. J., Yang, X. Y., et al. (2016). A study on criterion, identification and distribution of maturity groups for spring-sowing soybeans in Northeast China. *Soybean Sci.* 35, 181–192. doi: 10.11861/j.issn.1000-9841.2016.02.0181
- Gai, J., and Wang, Y. (2001). A study on the varietal eco-regions of soybeans in China. *Sci. Agric. Sin.* 34, 139–145.
- Gai, J., Xiong, D., and Zhao, T. (2015). *Pedigree and Germplasm Basis of Soybean Breeding Cultivars in China (1923–2005)*. Beijing: China Agricultural Press.
- Gai, J., Zhao, T., Cui, Z., and Qiu, J. (1998). Nuclear and cytoplasmic contributions of germplasm from distinct areas to the soybean cultivars released during 1923–1995 in China. *Sci. Agric. Sin.* 31, 35–43.
- Gizlice, Z., Carter, T. Jr., and Burton, J. (1994). Genetic base for North American public soybean cultivars released between 1947 and 1988. *Crop Sci.* 34, 1143–1151. doi: 10.2135/cropsci1994.0011183X003400050001x
- Goslee, S. C., and Urban, D. L. (2007). The ecodist package for dissimilarity-based analysis of ecological data. *J. Stat. Softw.* 22, 1–19. doi: 10.18637/jss.v022.i07
- Hallauer, A. R., and Miranda, J. B. (1981). *Quantitative Genetics in Maize Breeding*. Ames, IA: Iowa State Univ Press.
- He, J., Meng, S., Zhao, T., Xing, G., Yang, S., Li, Y., et al. (2017). An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theor. Appl. Genet.* 130, 2327–2343. doi: 10.1007/s00122-017-2962-9
- Hiromoto, D. (1986). The genetic base of Brazilian soybean (*Glycine max* L. Merrill) cultivars. *Rev. Bras. Genet.* 9, 295–306.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* 00, 1–3. doi: 10.48550/arXiv.1303.3997
- Li, S., Cao, Y., He, J., Zhao, T., and Gai, J. (2017). Detecting the QTL-allele system conferring flowering date in a nested association mapping population of soybean using a novel procedure. *Theor. Appl. Genet.* 130, 2297–2314. doi: 10.1007/s00122-017-2960-y
- Liu, F., He, J., Wang, W., Xing, G., Zhao, J., Li, Y., et al. (2021a). Genetic dynamics of flowering date evolved from later to earlier in annual wild and cultivated soybean in China. *Crop Sci.* 61, 2336–2354. doi: 10.1002/csc2.20462
- Liu, X., He, J., Wang, Y., Xing, G., Li, Y., Yang, S., et al. (2020). Geographic differentiation and phylogeographic relationships among world soybean populations. *Crop J.* 8, 260–272. doi: 10.1016/j.cj.2019.09.010
- Liu, X., Li, C., Cao, J., Zhang, X., Wang, C., He, J., et al. (2021b). Growth period QTL-allele constitution of global soybeans and its differential evolution changes in geographic adaptation versus maturity group extension. *Plant J.* 108, 1624–1643. doi: 10.1111/tpj.15531

FUNDING

This work was supported by the National Key Research Development Program of China (2021YFF1001204), the National Natural Science Foundation of China (32072082), the Program of Jiangsu Province (JBGS-2021-014), the MOE 111 Project (B08025), the China Agriculture Research System of MOF and MARA (CARS-04), the Primary Research & Development Plan of Jiangsu Province (BE2021358), the Jiangsu JCIC-MCP, the Cyrus Tang Innovation Center for Seed Industry (CTIC-SI201902) and the Guidance Foundation of Sanya Institute of Nanjing Agricultural University (NAUSY-ZZ02, NAUSY-MS05).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.945839/full#supplementary-material>

- Liu, X., Wu, J. A., Ren, H., Qi, Y., Li, C., Cao, J., et al. (2017). Genetic variation of world soybean maturity date and geographic distribution of maturity groups. *Breed. Sci.* 67, 221–232. doi: 10.1270/jsbbs.16167
- 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
- Meng, S., He, J., Zhao, T., Xing, G., Li, Y., Yang, S., et al. (2016). Detecting the QTL-allele system of seed isoflavone content in Chinese soybean landrace population for optimal cross design and gene system exploration. *Theor. Appl. Genet.* 129, 1557–1576. doi: 10.1007/s00122-016-2724-0
- Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326. doi: 10.1093/nar/8.19.4321
- 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
- Song, W., Sun, S., Ibrahim, S. E., Xu, Z., Wu, H., Hu, X., et al. (2019). Standard cultivar selection and digital quantification for precise classification of maturity groups in soybean. *Crop Sci.* 59, 1997–2006. doi: 10.2135/cropsci2019.02.0095
- Tang, F., Liang, J., Wei, Q., Chen, W., Guo, X., and Chen, Y. (2016). Study on maturity group classification of soybean varieties in Guangxi. *Soybean Sci.* 35, 891–895. doi: 10.11861/j.issn.1000-9841.2016.06.0891
- Weir, B. S., and Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* 38, 1358–1370. doi: 10.2307/2408641
- Wen, Z., Ding, Y., Zhao, T., and Gai, J. (2009). Genetic diversity and peculiarity of annual wild soybean (*G. soja* Sieb. et Zucc.) from various eco-regions in China. *Theor. Appl. Genet.* 119, 371–381. doi: 10.1007/s00122-009-1045-y
- Wysmierski, P. T., and Vello, N. A. (2013). The genetic base of Brazilian soybean cultivars: evolution over time and breeding implications. *Genet. Mol. Biol.* 36, 547–555. doi: 10.1590/S1415-47572013005000041
- Xiong, D., Zhao, T., and Gai, J. (2007). The core ancestors of soybean cultivars released during 1923–2005 in China. *Soybean Sci.* 26, 641–647.
- Xiong, D.-J., Zhao, T., and Gai, J. (2008). Geographical sources of germplasm and their nuclear and cytoplasmic contribution to soybean cultivars released during 1923 to 2005 in China. *Acta Agron. Sin.* 34, 175–183. doi: 10.3724/sp.J.1006.2008.00175
- Zhang, Y., He, J., Wang, Y., Xing, G., Zhao, J., Li, Y., et al. (2015). Establishment of a 100-seed weight quantitative trait locus-allele matrix of the germplasm population for optimal recombination design in soybean breeding programmes. *J. Exp. Bot.* 66, 6311–6325. doi: 10.1093/jxb/erv342
- Zhao, T., and Gai, J. (2004). The origin and evolution of cultivated soybean [*Glycine max* (L.) Merr.]. *Sci. Agric. Sin.* 37, 954–962.
- Zhou, X., Carter, T. E., Cui, Z., Miyazaki, S., and Burton, J. W. (2000). Genetic base of Japanese soybean cultivars released during 1950–1988. *Crop Sci.* 40, 1794–1802. doi: 10.2135/cropsci2000.4061794x

Conflict of Interest: 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Li, Wang, Pan, Liu, He, Liu, Cao, Zhang, Zhao and Gai. 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(s) 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

Xia Xin,
Institute of Crop Sciences (CAAS), China

REVIEWED BY

Zhoufei Wang,
South China Agricultural University, China
Baoshen Liu,
Shandong Agricultural University,
China

*CORRESPONDENCE

Zhenhua Wang
zhenhuawang_2006@163.com
Hong Di
dihongdh@163.com

[†]These authors have contributed equally to this work

SPECIALTY SECTION

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

RECEIVED 08 June 2022

ACCEPTED 24 June 2022

PUBLISHED 18 July 2022

CITATION

Li C, Jia Y, Zhou R, Liu L, Cao M, Zhou Y,
Wang Z and Di H (2022) GWAS and
RNA-seq analysis uncover candidate genes
associated with alkaline stress tolerance in
maize (*Zea mays* L.) seedlings.
Front. Plant Sci. 13:963874.
doi: 10.3389/fpls.2022.963874

COPYRIGHT

© 2022 Li, Jia, Zhou, Liu, Cao, Zhou, Wang
and Di. 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(s) 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.

GWAS and RNA-seq analysis uncover candidate genes associated with alkaline stress tolerance in maize (*Zea mays* L.) seedlings

Chunxiang Li[†], Yue Jia[†], Runyu Zhou, Liwei Liu, Mengna Cao,
Yu Zhou, Zhenhua Wang* and Hong Di*

Key Laboratory of Germplasm Enhancement, Physiology and Ecology of Food Crops in Cold Region, Ministry of Education, Northeast Agricultural University, Harbin, China

Soil salt-alkalization is a common yet critical environmental stress factor for plant growth and development. Discovering and exploiting genes associated with alkaline tolerance in maize (*Zea mays* L.) is helpful for improving alkaline resistance. Here, an association panel consisting of 200 maize lines was used to identify the genetic loci responsible for alkaline tolerance-related traits in maize seedlings. A total of nine single-nucleotide polymorphisms (SNPs) and their associated candidate genes were found to be significantly associated with alkaline tolerance using a genome-wide association study (GWAS). An additional 200 genes were identified when the screen was extended to include a linkage disequilibrium (LD) decay distance of $r^2 \geq 0.2$ from the SNPs. RNA-sequencing (RNA-seq) analysis was then conducted to confirm the linkage between the candidate genes and alkali tolerance. From these data, a total of five differentially expressed genes (DEGs; $|\log_2FC| \geq 0.585$, $p < 0.05$) were verified as the hub genes involved in alkaline tolerance. Subsequently, two candidate genes, *Zm00001d038250* and *Zm00001d001960*, were verified to affect the alkaline tolerance of maize seedlings by qRT-PCR analysis. These genes were putatively involved protein binding and “flavonoid biosynthesis process,” respectively, based on Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) enrichment analyses. Gene promoter region contains elements related to stress and metabolism. The results of this study will help further elucidate the mechanisms of alkaline tolerance in maize, which will provide the groundwork for future breeding projects.

KEYWORDS

maize, alkali tolerance, genome-wide association study, RNA-seq, candidate genes

Introduction

Soil salt-alkalization is a major source of stress to plants, and therefore a threat to agricultural development and crop productivity (Zhu, 2016). Due to the increase in global temperature caused by human activity and industrial development, the serious evaporation of soil water, and the lack of drainage consideration during irrigation has led to the

deepening of land salinization, threatening crop yields. The effect of salty soils on plants includes salt stress caused by neutral salt (NaCl and Na₂SO₄; Xu et al., 2019), alkaline salt stress caused by alkaline salt (NaHCO₃ and Na₂CO₃; Yang et al., 2019), and saline-alkali stress caused by both alkaline and neutral salt (Shi and Sheng, 2005). Several studies have shown that alkaline salts, due to their high pH, are more dangerous to plants than neutral salts (Campbell and Nishio, 2000; Hartung et al., 2002; Chen et al., 2012).

Saline-alkali stress induces a variety of plant responses, including changes in plants' physiological, biochemical, and morphological structures. It also causes ionic imbalance and changes in osmotic pressure (100 mM NaCl; Chaparzadeh et al., 2004; Ambede et al., 2012; Abreu et al., 2013; Mishra et al., 2013). For example, stress accumulation of Na⁺ leads to competition with K⁺ for protein binding, which inhibits protein synthesis (Schachtman and Liu, 1999; Pardo and Quintero, 2002). Physiologically speaking, salt stress reduced stomatal number, stomatal density, and photosynthesis, and increased the respiration rate in plants (100 and 200 mM NaCl) (Dikobe et al., 2021). Biochemically, enzyme content such as amount of soluble proteins, total free amino acids, prolines, Na⁺, and malondialdehyde was increased under alkaline treatment (75 mM Na₂CO₃) (Abdel Latif and Tran, 2016). Saline soil also leads to decreased seed vigor, reduced germination, damaged root cell structure, reduced nutrient absorption and utilization, and ultimately reduced yield as a culmination of all these factors (mixed salts of NaCl, Na₂SO₄, NaHCO₃, and Na₂CO₃; Peng et al., 2008; Gao et al., 2014). For example, under different concentrations of saline irrigation, maize yield decreased by 2.08%–3.93% with an increase of the salt concentration (NaCl; Feng et al., 2017). Soil salinization affected the physiological and biochemical indicators, or other resistance, ultimately leading to yield impacts.

Maize (*Zea mays* L.) is an important cereal crop in the world. Among the cereal crops, it is the most sensitive to alkalinity. Therefore, exploiting genes associated with alkaline tolerance is helpful for improving alkaline resistance in this crop plant. However, not many genes have been associated with alkaline tolerance. Currently, there has been some progress in identifying the genetic loci associated with these tolerances in crop plants. For instance, a major quantitative trait locus (QTL) for photosynthetic alkali tolerance was detected for two consecutive years by the RIL population in rice (8.5 mM Na₂CO₃; Sun et al., 2019). Three pairs of additive by additive (AA) epistatic QTLs associated with dead leaf rate under alkali tolerance were also identified (pH 8.7 to 8.9, natural alkaline soil; Liang et al., 2014). In maize, researchers identified two QTLs associated with alkali tolerance by using simple sequence repeats (SSR) and specific locus amplified fragment-sequencing (SLAF-seq) markers through 151 F_{2:3} populations (100 mM Na₂CO₃; Zhang et al., 2018). For example, the researchers mapped a major QTL for alkali tolerance on Chr 2 in maize by identifying sodium ion content and validating the candidate gene (100 mM NaHCO₃; Cao et al., 2020). When maize was

treated with alkaline salt, the ZmNSA1 protein bound to Ca²⁺ for degradation and made H⁺ efflux, ultimately promoting saline-alkaline tolerance in the treated plants (100 mM NaHCO₃ or NaCl; Cao et al., 2020).

In addition to using QTLs to discover alkaline stress-related traits, GWAS have also been used to detect several SNPs associated with this trait. GWAS has been shown to be an effective method to identify genes and alleles associated with certain agronomic traits under complex environments (Yan et al., 2011; Li et al., 2016). Depending on the rapid decay of LD and amount of diversity, GWAS provides a systematic approach to the analysis of complex quantitative traits in many crops, including maize (Li et al., 2013). In the whole wheat genome, 326,570 SNPs were used in the SNP-GWAS method to unearth 20 SNPs that were significantly associated with grain length and 31 SNPs with grain width, among which the IWB32119 marker located on chromosome 2A was associated with the grain weight locus gene *TaFlo2-A1* (Sajjad et al., 2017; Li et al., 2019). As a benefit of the rapid decay of maize LD, numerous loci controlling complex traits have been identified in maize by GWAS. Researchers extracted two candidate genes associated with low-temperature tolerance in maize by GWAS mining and used RNA-seq data analysis to confirm that these genes were in fact related to low-temperature tolerance (Zhang et al., 2020). Seven candidate genes associated with drought were identified by combining GWAS, DEGs, and co-expression analysis (Guo et al., 2020). Development of the functional SNP marker Sh2_rs844805326 for the prediction of sweet corn was achieved using GWAS (Ruanjaichon et al., 2021). GWAS has also been used to mine genes that control complex traits in maize, such as leaf development, plant height, and ear height (Li et al., 2016; Miculan et al., 2021). As useful as the GWAS method is, there are few reports on the mining of maize alkali tolerance-related genes using GWAS.

In this study, a panel of 200 different inbred maize lines was used to analyze ten traits associated with alkali resistance by GWAS. Candidate genes were identified and further confirmed using RNA-seq data find the following: (a) SNPs associated with resistance to alkaline, (b) alkali tolerance maize inbred lines, and (c) relevant candidate genes for future agricultural research and breeding applications.

Materials and methods

Plant materials

A total of 200 inbred maize lines with extensive variation in yield traits and biotic stress tolerances were analyzed in this study (Supplementary Table S1; Weng et al., 2011; Liu et al., 2015; Zhang et al., 2017). The seeds were planted in Harbin, Heilongjiang Province, and produced by manual self-pollination. The field site used was well-managed and free from pests and diseases. After seeds were harvested and dried completely, they were stored in a

4°C seed cabinet until ready for use. The 200 maize inbred lines had a >90% germination rate at 25°C, as shown by a previous study conducted. The alkaline-resistant maize line K10 was analyzed using RNA-seq to assess the expression levels of the whole genome.

Seed germination and alkaline stress treatment during seedling stage

To germinate the maize seeds, they were first sterilized in 10g/l sodium hypochlorite for 20 min. After rinsing the seeds three times with sterilized distilled water, they were imbibed for 6 h at 25°C, then the seeds were subjected to the standard germination: briefly, 30 sterile seeds were placed between two wet germination papers (Li et al., 2018). Seed germination was defined as the observation of a 0.5 cm radicle emergence. After 48 h of seed germination, the seedlings with the same bud length were selected and transferred to pots to continue to grow, with 5 seedlings per pot and 30 seedlings per treatment. Seeds were grown in the growth chamber with an 18 h, 25°C/6 h, 22°C light/dark, and temperature/light cycle (Ma et al., 2020). When the seedlings reached the three-leaf growth stage, they were treated with ½ Hoagland solution containing 25 mmol/l Na₂CO₃ every day for 10 days (Geng et al., 2020). The control plants were also irrigated with ½ Hoagland solution. The treatment group was watered with ½ Hoagland every 30 days to prevent the excessive accumulation of Na₂CO₃. Three independent experiments were used for each treatment and the control.

Phenotypic data measurement and analysis

After Na₂CO₃ treatment for 10 days, maize roots were collected, rinsed with deionized water, and dried by careful blotting before measurements were taken. For root trait data analysis, the Epson Perfection V800 scanner was used to measure root traits to be analyzed by the Regent WinRHIZO (Canada) software. The root traits were root length (RL), root average diameter (RAD), root surface area (RSA), root volume (RV), and root tip number (RTN). Root fresh weight (RFW) and shoot fresh weight (SFW) were measured immediately after observing no moisture in the seedlings. After drying at 105°C for 30 min, roots were transferred to 80°C to continue to dry until they reached a constant weight, then the root dry weight (RDW) and shoot dry weight (SDW) were measured. A ruler was used to measure seed length (SL). The relative performance of the 10 traits was used as an indicator of alkali tolerance. Different treatments consisted of 30 seedlings. Three seedlings were measured for each trait, and the average value of these measurements was taken.

Phenotypic data for the mean, maximum, minimum, kurtosis, skewedness, variance analysis, and standard deviation for each

relative trait were analyzed by IBM SPSS Statistics (20.0).¹ Correlation analysis of phenotype data was performed with the “Performance Analytics” package in R and IBM SPSS Statistics (20.0). ANOVA for 10 alkali tolerance-related traits was calculated using IBM SPSS Statistics (20.0). The broad-sense heritability (H^2) was estimated using the following formula: $H^2 = V_G / (V_G + V_E)$, with H^2 as the heritability estimate, V_G as the genotype variance, and V_E as the environmental variance.

Genotypic data and GWAS analysis

Using the Illumina BeadStation 500 G SNP developed by Illumina, 41,110 SNPs were selected from 56,110 SNPs with a minor allele frequency (MAF) > 0.05. SNPs with < 20% heterozygosity were used for analysis (Weng et al., 2011; Liu et al., 2015; Zhang et al., 2017). From this, a total of 40,697 SNPs were given for association analysis. 7,742 distributed SNP datasets were assessed for structural parameters by using the STRUCTURE 2.3 software (Evanno et al., 2005). ΔK was calculated using Structure Harvester (Earl and vonHoldt, 2011). Using the software TASSEL 5.0, the kinship information for the 200 inbred lines was estimated.

Using the MLM model in TASSEL 5.0 (Bradbury et al., 2007), the association analysis was performed on the 10 relative traits measured at the seedling stage, using 0.05/Ne to calculate the threshold for association analysis. Because the Bonferroni correction ($0.05/41,110 = 1.22 \times 10^{-6}$) was too conservative, 10 trait-associated loci were considered rare; therefore, the use of a less stringent threshold of $-\log_{10}(P) > 4$ was applied for detecting significant associations (Weng et al., 2011; Zhang et al., 2017; Jia et al., 2020). The Manhattan plot was then generated with the CMplot package in R. When $r^2 = 0.1$, the LD decay was 55, 60, 110, 100, 60, 45, 80, 70, 60, and 110 across chromosomes 1–10, respectively. At $r^2 = 0.2$, the LD decay was 395, 520, 710, 610, 770, 575, 775, 775, 1,125, and 850 across chromosomes 1–10, respectively (Zhang et al., 2020). Significantly associated SNPs with a physical distance less than the LD decay distance that were also located on the same chromosome were defined as one locus. The upstream and downstream ranges of the corresponding LD decay distances for each locus were used to further mining genes. The genetic information of candidate genes was retrieved using the gene model from the B73 RefGen_V4 in the MaizeGDB website.²

Transcriptomic analysis

The maize inbred line K10, which had high alkaline tolerance, was used for RNA-seq analysis. Before germination, the seeds were treated with the same sterilization method described in

¹ <https://www.ibm.com/support/pages/node/230551>

² <http://www.maizegdb.org/>

section Seed Germination and Alkaline Stress Treatment During Seedling Stage. The treated and control groups were sampled after 3 and 4 days Na_2CO_3 treatment, at the three-leaf stage. Leaves were taken from each plant, frozen in liquid nitrogen, and cryopreserved at -80°C until RNA extraction with Trizol. The maize RNA was sent to DATA Corporation (Beijing, China) for library construction. Genomic DNA was removed using DNase I and mRNA was enriched using oligo (dT) magnetic beads and subsequently fragmented. The fragmented products were then amplified by PCR and sequenced after the samples passed the in-house quality inspection.

For our data analysis, the maize genome ZmB73RefGenV4³ was used as the reference genome. Gene expression was calculated using Cuffdiff software and expressed as fragments per kilobase of transcript per million fragments mapped (FPKM) value. The threshold for the significant differential expression of genes was set to $p < 0.05$ and $|\log_2\text{FC}| \geq 0.585$.

Identification and annotation of candidate genes

GO enrichment analysis was used to identify all GO terms for DEGs that were significantly enriched in background genes. All DEGs utilized publicly available databases⁴ for GO term classification and grouping. A hyper-geometric test and Phytozome⁵ were used to identify significantly enriched GO terms. GO terms were divided into three categories: “biological process,” “cellular component,” and “molecular function” (Du et al., 2010). To further understand the biological functions of DEGs, enrichment analysis was performed using the KEGG pathway⁶ (Kanehisa et al., 2008). Genes were annotated using the MaizeGDB and NCBI⁷ databases. Heatmaps were plotted using an online site.⁸

Quantitative real-time PCR analysis

A qRT-PCR assay was performed using the same total RNA used for the transcriptome data analysis of the alkaline-tolerant inbred line K10 and the alkaline-sensitive inbred line Mo17. For each sample, the *TransScript*[®] One-Step gDNA Removal Kit and cDNA Synthesis SuperMix (TransGen Biotech, Beijing, China) was used to synthesize cDNA. The real-time PCR System (Analytik Jena, Germany) was used for qRT-PCR analysis. As a control, actin was used to normalize the obtained Ct values.

Primer5.0 was used to design specific primers for this assay (Supplementary Table S2). The 20 μl reaction mixtures consisted of 2 μl cDNA, 7.2 μl of ddH₂O, 0.4 μl each of the forward and reverse primers (10 μM), and 10 μl of $2 \times \text{TransStart}^{\text{®}}$ Tip Green qPCR SuperMix (TransGen Biotech, Beijing, China). Relative gene expression was calculated using the $2^{-\Delta\Delta\text{CT}}$ analytical method (Livak and Schmittgen, 2001). Three biological replicates were combined to be used for three technical replicates, and the average of the three replicates was used for gene expression analysis.

Results

Phenotypic traits related to alkaline tolerance

The alkaline tolerance of 200 inbred lines at the seedling stage was evaluated (Figure 1). The descriptive statistics for alkaline resistant traits RSL, RSFW, RRFW, RSDW, RRDW, RRL, RRAD, RRSA, RRV, and RRTN evaluated in an artificial environment are presented in Table 1. Under both the control and alkaline conditions, the indicators of alkali tolerance were determined, and descriptive statistics were executed for the related characters, including the determination of the maximum value, minimum value, and standard deviation (SD). For these traits, there were abundant phenotypic variations, they ranged from 0.5487 (RSFW) to 1.0847 (RRAD). The RRL varied from 0.2908 to 0.9935, the RRTN varied from 0.1959 to 0.9916, the RSDW varied from 0.2846 to 0.9972, the RSL varied from 0.4709 to 0.9830, the RSFW varied from 0.1857 to 0.9399, the RRFW varied from 0.4069 to 0.9958, the RRDW varied from 0.4448 to 0.9957, the RRSA varied from 0.3679 to 0.9986, the RRV varied from 0.4152 to 0.9971, and the RRAD varied from 0.8180 to 1.4405. Curiously, the RRAD of some inbred lines had a value of >1.000 , suggesting that the relative root average diameter of these inbred lines was actually stimulated by the alkali treatment. In general, the 10 traits analyzed in the panel fit a normal distribution. ANOVA tests showed statistically significant genotypic differences for the 10 relative traits. In the association panel, the broad-sense heritability (H^2) of the 10 relative traits ranged from 24.00% (RRAD) to 67.30% (RRTN; Supplementary Table S3).

In total, 10 traits were measured (Figure 2) and their correlations with each other were determined using the relative value of each trait under alkali treatment. The traits analyzed at the seedling stage characterized maize growth, indicating that these ten traits can be used as indicators of maize alkali tolerance at the seedling stage. Except for in the case of RRV and RRAD, RSFW and RRAD, RRDW and RRAD, RRAD and the other traits had a significant negative correlation ($p < 0.05$), such that RRL and RRAD were -0.49 , the RRTN and RRAD were -0.42 , and the RSL and RRAD were -0.27 ($p < 0.001$). Except for the RRAD, significant positive correlations were achieved between the other traits ($r = 0.17\text{--}0.87$, $p < 0.05$). The three traits that were most significantly associated were RRL and RRSA (0.87), RRV and

³ https://www.ncbi.nlm.nih.gov/data-hub/genome/GCA_000005005.6/

⁴ <http://www.geneontology.org/>

⁵ <https://phytozome-next.jgi.doe.gov/>

⁶ <http://www.genome.ad.jp/kegg/kegg2.html>

⁷ <http://www.ncbi.nlm.nih.gov/>

⁸ <http://www.bioinformatics.com.cn>

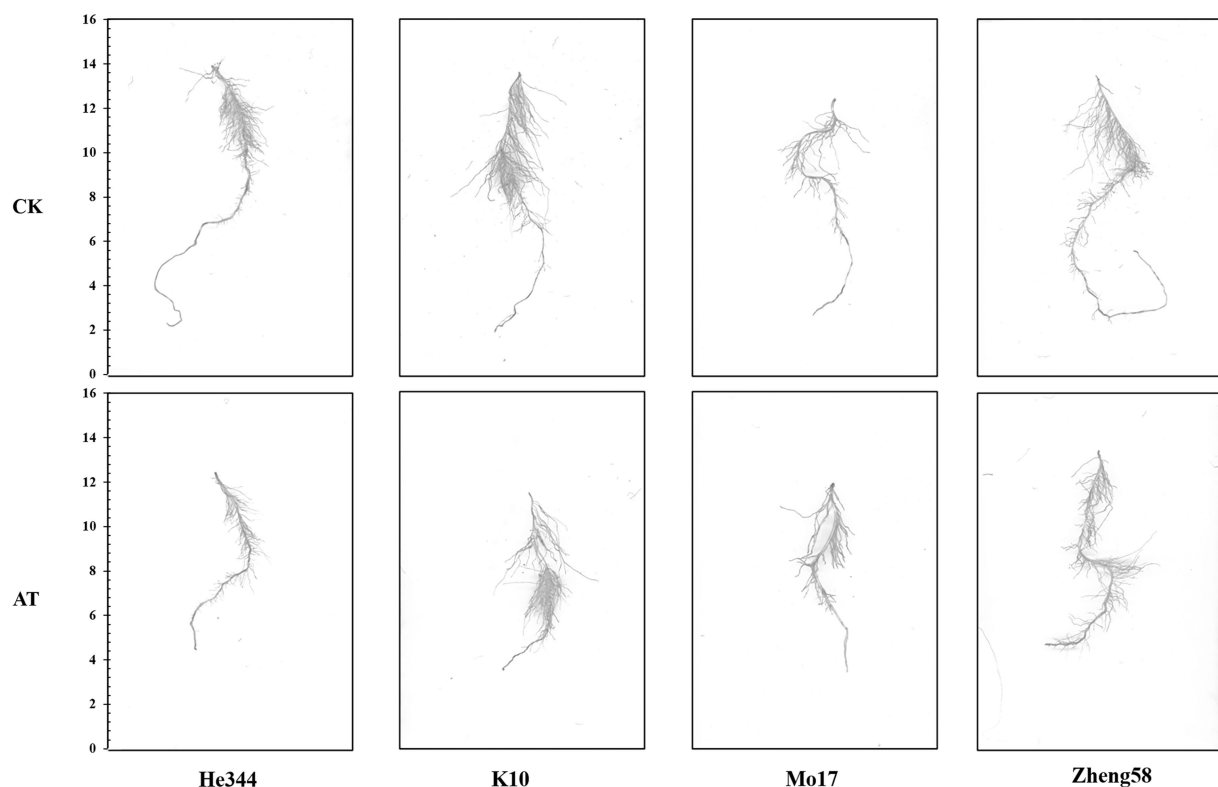


FIGURE 1

Morphological change in maize under control (CK) and alkaline (AT) treatment. Comparison of roots of maize inbred line He344; inbred line k10; inbred line Mo17; and inbred line Zheng58 under control and alkaline treatment for 10 days. The alkali-sensitive inbred lines Mo17 and He344 are from the Lancaster group, and the alkali-tolerant inbred lines Zheng58 and K10 are from the PA group.

RRSA (0.81), RRL and RRTN (0.71), and RSL and RSDW (0.71; $p < 0.001$; [Supplementary Table S4](#)). The 10 traits were approximated as a normal distribution (kurtosis and skewness $< \pm 1$), and showed a significant correlation, indicating the suitability of the 10 traits for further GWAS analysis.

Population structure and GWAS analysis

The population structure was calculated using STRUCTURE version 2.3. When $K = 6$, the 200 inbred lines could be grouped into six large subgroups ([Supplementary Figure S1](#)). GWAS was used for the 10 relative traits (RSL, RSFW, RRFW, RSDW, RRDW, RRL, RRAD, RRSA, RRV, and RRTN) for the 200 maize inbred lines. GWAS of the SNP markers and related traits was performed using TASSEL 5.0 software, using a mixed linear model (MLM) combined with population structure and kinship. From this analysis, nine SNPs were shown to be significantly associated with four traits (the p values ranged from $8.1081\text{E-}06$ to $9.6125\text{E-}05$; [Figure 3](#)).

Among the nine associated SNPs, the distance between bt2.5 and bt2.8 on chromosome 4 was only 706 bp (for RRDW), and the distance between PZE-103084802 and PZE-103084794 on chromosome 3 was only 42 kb (for RRAD). These SNPs were

located on maize chromosomes 3, 4, 5, 6, and 9, with the highest number of SNPs (3) on chromosome 4. A single SNP associated with RRT was discovered. Two SNPs were detected for RSL, and the most significantly associated SNP was for PZE-106099144 ($p = 8.1081\text{E-}06$). The three SNPs were significantly associated with RRAD and were located on chromosomes 4 (SYN9460, $p = 7.21\text{E-}05$), 3 (PZE-103084802, $p = 7.45\text{E-}05$), and 3 (PZE-103084794, $p = 8.83\text{E-}05$). The three SNPs were significantly associated with RRDW and were located on chromosomes 9 (PZE-109058967, $p = 9.18\text{E-}05$), 4 (bt2.5, $p = 9.28\text{E-}05$), and 4 (bt2.8, $p = 9.28\text{E-}05$). Four relative traits containing 9 SNPs explained 7.94%–10.71% of the total phenotypic variation ([Table 2](#)).

Among the nine SNPs found for the four traits, nine candidate genes that were directly closest to the physical location of the SNPs were mined using the B73 RefGen_v4 Maize Gene Database ([Table 3](#)). Two genes were found to be associated with RSL (*Zm00001d038265*, and *Zm00001d038320*). Three genes were found to be associated with RRAD (*Zm00001d050905*, *Zm00001d041766*, and *Zm00001d041767*). Three genes were found to be associated with RRDW (*Zm00001d046591*, *Zm00001d050109*, and *Zm00001d050110*). For RRTN, only one candidate gene was associated (*Zm00001d013802*). Furthermore, when the LD decay distance was changed to $r^2 = 0.2$, an additional

TABLE 1 Descriptive statistics of seedling stage traits under alkaline and normal conditions.

Trait	<i>n</i>	Mean	Maximum	Minimum	SD	Kurtosis	Skewness
RSL	200	0.7277	0.9830	0.4709	0.1013	−0.4045	0.0909
RSFW	200	0.5487	0.9399	0.1857	0.1476	−0.5848	0.1255
RRFW	200	0.7641	0.9958	0.4069	0.1270	−0.5716	0.0524
RSDW	200	0.6412	0.9972	0.2846	0.1524	−0.3523	−0.5156
RRDW	200	0.8274	0.9957	0.4448	0.1104	0.0542	−0.6983
RRL	200	0.6664	0.9935	0.2908	0.1610	−0.7076	−0.1055
RRAD	200	1.0847	1.4405	0.8180	0.1030	0.1840	0.2458
RRSA	200	0.7116	0.9986	0.3679	0.1438	−0.8569	−0.0881
RRV	200	0.7448	0.9971	0.4152	0.1405	−0.8175	−0.2192
RRTN	200	0.6806	0.9916	0.1959	0.1876	−0.7629	−0.2837

The traits of 200 inbred lines were evaluated at the seedling stage under normal and alkaline conditions (25 mmol/LNa₂CO₃). Ten relative traits were obtained by dividing the directly measured trait values under alkaline conditions (RL, RAD, RSA, RRV, RTN, SDW, RDW, SFW, and RFW) by the corresponding values under normal conditions.

200 candidate genes were added to the 9 associated SNPs (Supplementary Table S5). They were distributed on chromosomes 1–10: the PZE-106099144 (SL) locus contained 45 genes, the SYN24465 (SL) locus contained 42 genes, the SYN9460 (RAD) locus contained 10 genes, the PZE-103084802 (RAD) locus contained 17 genes, the PZE-103084794 (RAD) locus contained 19 genes, the PZE-109058967 (RDW) locus contained 36 genes, the bt2.5 (RDW) locus contained 9 genes, the bt2.8 (RDW) locus contained 9 genes, and the SYN30436 (RRTN) locus contained 39 genes.

RNA-seq analysis of leaves transcripts in response to alkali stress

To help identify candidate genes, the maize inbred line K10 was selected as an alkali stress tolerant example and its genome-wide gene expression levels were analyzed using RNA-seq. For downregulated DEGs, two were identified in the K4D_ATvsK4D_CK comparison group and 141 DEGs were identified in the K3D_ATvsK3D_CK comparison group. For the upregulated DEGs ($|\log_2FC| \geq 0.585$, $p < 0.05$), 17 were identified in the K4D_ATvsK4D_CK comparison group, 414 were identified in the K3D_ATvsK3D_CK comparison group, and four were identified in the K4D_ATvsK4D_CK and K3D_ATvsK3D_CK comparison groups (Figure 4).

A total of 209 candidate genes were associated with SNPs, and one important DEG was identified, including one candidate gene (*Zm00001d038250*) that was only upregulated in the K3D_ATvsK3D_CK comparison group. Four DEGs were upregulated in the K3D_ATvsK3D_CK and K4D_ATvsK4D_CK comparison group, including four candidate genes (*Zm00001d027619*, *Zm00001d032973*, *Zm00001d001820*, and *Zm00001d001960*). Putative protein functions are listed in Table 4, determined by NCBI. *Zm00001d038250* encoded a HSP40/DNAJ peptide-binding protein, *Zm00001d027619* encoded a beta-amylase, *Zm00001d032973* encoded a glycerol-3-phosphate acyltransferase7, *Zm00001d001820* encoded a protochlorophyllide

reductase 1, and *Zm00001d001960* encoded a flavanone 3-dioxygenase 1.

Functional prediction of candidate genes by GWAS and RNA-seq analysis

There were a total of 37 GO terms associated with the five candidate genes, three candidate genes were associated with four KEGG pathways, and one KEGG pathway (*zma01110*) was co-enriched by the three genes (Tables 5, 6). These GO terms were functionally related to three broad categories. The first function described was cellular components, including cytoplasmic part (*Zm00001d001820* and *Zm00001d038250*), cytoplasm (*Zm00001d001820* and *Zm00001d038250*), cell part (*Zm00001d038250* and *Zm00001d001820*), cell (*Zm00001d001820* and *Zm00001d038250*), cytosol (*Zm00001d038250*), intracellular part (*Zm00001d038250* and *Zm00001d001820*), intracellular (*Zm00001d038250*), membrane (*Zm00001d032973* and *Zm00001d001820*), plastid (*Zm00001d001820*), membrane-bounded organelle (*Zm00001d001820*), intracellular membrane-bounded organelle (*Zm00001d001820*), intracellular organelle (*Zm00001d001820*), and organelle (*Zm00001d001820*). The second type described was biological processes, which involved cellular process (*Zm00001d038250* and *Zm00001d032973*), metabolic process (*Zm00001d027619*, *Zm00001d032973*, *Zm00001d001820*, and *Zm00001d001960*), catabolic process (*Zm00001d027619*), carbohydrate metabolic process (*Zm00001d027619*), primary metabolic process (*Zm00001d027619*), organic substance metabolic process (*Zm00001d027619*), biosynthetic process (*Zm00001d032973*, *Zm00001d001820* and *Zm00001d001960*), response to abiotic stimulus (*Zm00001d001960*), response to stimulus (*Zm00001d001960*), photosynthesis (*Zm00001d001820*), and cellular metabolic process (*Zm00001d001820*). The third type of function described was molecular functions, which involved binding (*Zm00001d001960* and *Zm00001d038250*), protein binding (*Zm00001d038250*), catalytic activity (*Zm00001d027619*, *Zm00001d032973*, *Zm00001d001820*, and *Zm00001d001960*),

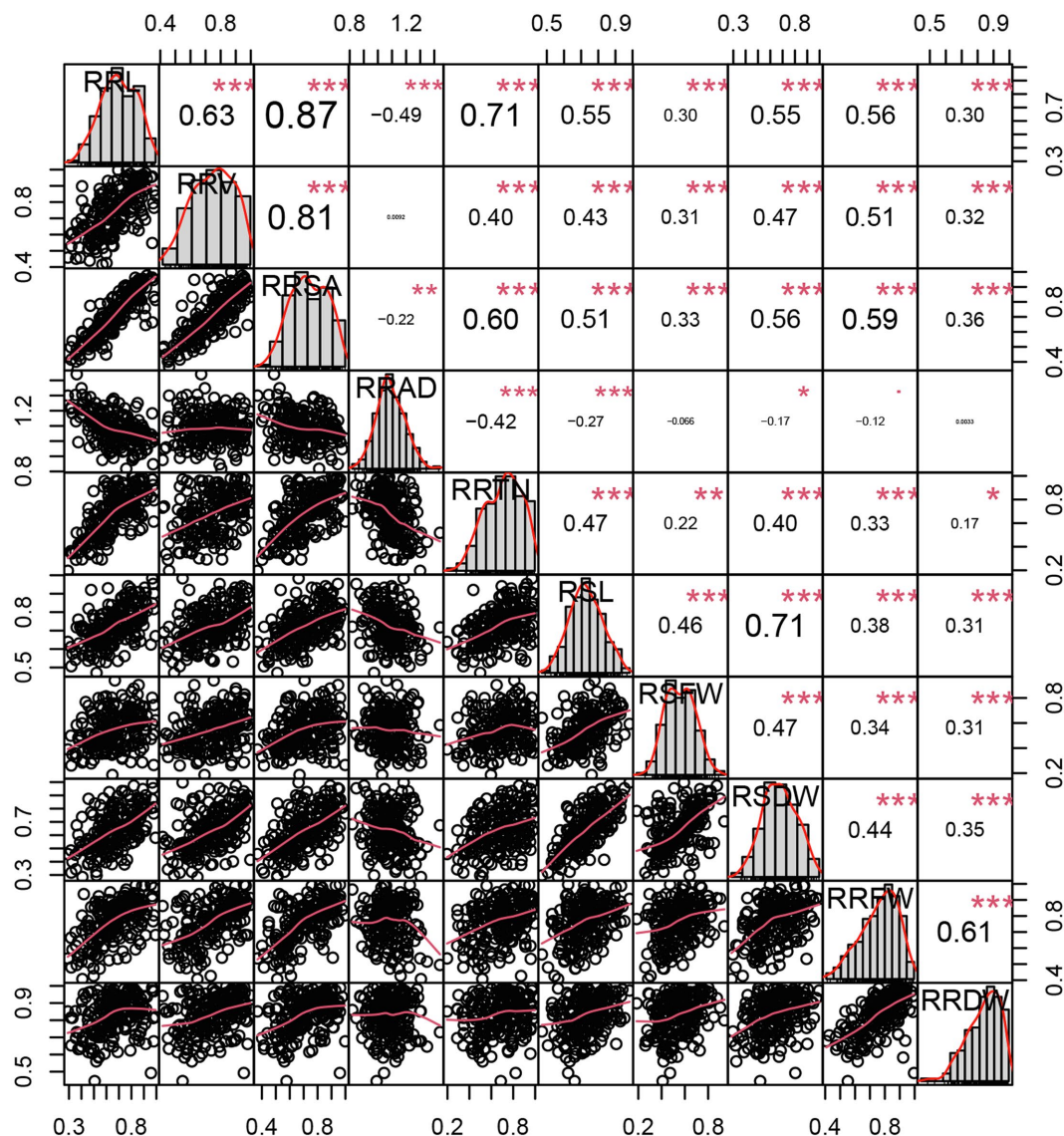


FIGURE 2

Distributions of and correlations between 10 relative phenotypic traits. The frequency distribution histograms of 10 traits are located on the diagonal line, the area below the diagonal line is the scatter plot of the traits, and the area above is the correlation coefficient between each pair of traits. *, **, and *** indicate significance at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

hydrolase activity (*Zm00001d027619* and *Zm00001d032973*), and transferase activity (*Zm00001d032973*; Table 5; Figures 5A,B). Two major classes of KEGG pathways were identified: the flavonoid synthesis pathway and the metabolic pathway, including starch and sucrose metabolism, and porphyrin and chlorophyll metabolism (Table 6; Figure 5C).

qRT-PCR validation for differentially expressed genes

Five candidate genes were selected based on RNA-seq ($|\log_2FC| \geq 0.585$, $p < 0.05$) for qRT-PCR analysis:

Zm00001d038250, *Zm00001d027619*, *Zm00001d032973*, *Zm00001d001820*, and *Zm00001d001960*. When alkali treated for 3 days, the expression level of one gene (*Zm00001d038250*) in K10 was significantly higher (1.49) than that in Mo17 (-0.74) ($p < 0.01$). The expression level of this gene showed a downward trend after 4 days of treatment, but the expression level in B73 (0.410) was still higher than that in Mo17 (-0.237). The expression level of gene *Zm00001d027619* in K10 (0.263) was significantly lower than that in Mo17 (2.140) when treated for 3 days, and there was no significant difference when treated for 4 days. The expression levels of the other two genes (*Zm00001d032973* and *Zm00001d001820*) in K10 (2.617 and 1.787) were significantly higher than those in Mo17

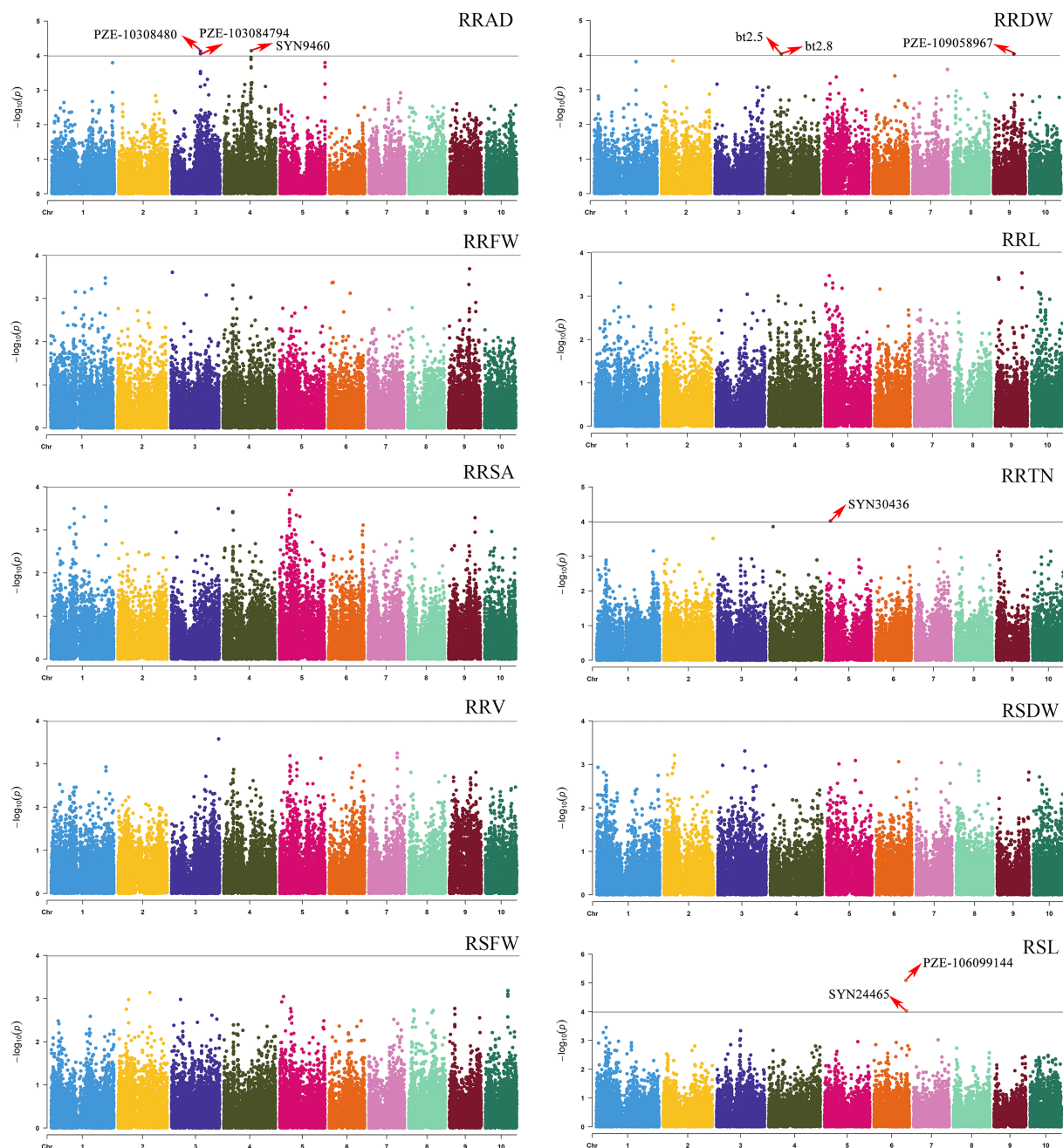


FIGURE 3

Manhattan plots of GWAS results showing the significant SNPs associated with 10 relative phenotypic traits. The traits included the following: the relative root average diameter (RRAD), relative root fresh weight (RRFW), relative root surface area (RRSA), relative root volume (RRV), relative shoot fresh weight (RSFW), relative root dry weight (RRDW), relative root length (RRL), relative root tip number (RRTN), relative shoot dry weight (RSDW), and relative seedling length (RSL).

(1.610 and -3.867) when treated with alkali for 3 days but showed the opposite pattern when treated for 4 days: the expression level in Mo17 (3.533 and 3.353) was significantly higher than that in K10 (0.233 and 0.337). The expression level of the last gene (*Zm00001d001960*) in K10 (2.933 and 0.960) was significantly higher than that of Mo17 (-0.593 and -1.130) at 3 and 4 days of alkali treatment (Figure 6). The

results of RNA-seq and qRT-PCR analysis showed that the expression trends of the five genes were consistent: the expressions were all upregulated after 3 or 4 days of alkaline treatment.

Two candidate genes were shown to be more attractive than the others: *Zm00001d038250* and *Zm00001d001960*, as they were upregulated in K10 and downregulated in Mo17 under both the

TABLE 2 SNPs associated with alkaline tolerance detected in 200 maize inbred lines.

Trait	SNP	Alleles	Chromosome	Position	<i>p</i> -value	<i>R</i> ²
RSL	PZE-106099144	A/C/G/T	6	152,684,808	8.11E-06	10.5%
RRAD	SYN9460	A/G	4	129,386,027	7.21E-05	10.7%
RRAD	PZE-103084802	A/C/T	3	136,692,058	7.45E-05	8.7%
RRAD	PZE-103084794	C/G/T	3	136,649,558	8.83E-05	8.7%
RRDW	PZE-109058967	A/G	9	97,545,021	9.18E-05	8.4%
RRDW	bt2.5	A/G	4	66,290,049	9.28E-05	8.3%
RRDW	bt2.8	A/G	4	66,290,755	9.28E-05	8.3%
RSL	SYN24465	C/G/T	6	154,387,810	9.35E-05	7.9%
RRTN	SYN30436	A/G	5	20,797,115	9.61E-05	9.3%

The location is indicated by the chromosome and base pair position. Values of *p* less than 10^{-4} corresponding to a 5% type I error are displayed in scientific notation. *R*²: the percentage of phenotypic variation.

TABLE 3 Functions of candidate genes that are directly associated with alkaline tolerance traits.

Trait	SNP	Gene ID	Gene function
RSL	PZE-106099144	<i>Zm00001d038265</i>	uncharacterized
RRAD	SYN9460	<i>Zm00001d050905</i>	dynein light chain type 1 family protein
RRAD	PZE-103084802	<i>Zm00001d041767</i>	PHD finger protein
RRAD	PZE-103084794	<i>Zm00001d041766</i>	transcription initiation factor TFIID subunit 8
RRDW	PZE-109058967	<i>Zm00001d046591</i>	vacuolar-type H ⁺ -pyrophosphatase5
RRDW	bt2.5	<i>Zm00001d050109</i>	water channel
RRDW	bt2.8	<i>Zm00001d050110</i>	aspartic proteinase nepenthesin-1
RSL	SYN24465	<i>Zm00001d038320</i>	ethylene-responsive transcription factor ABR1
RRTN	SYN30436	<i>Zm00001d013802</i>	uncharacterized

treatment durations (3 or 4 days). Gene function annotation suggested that they may be involved in alkaline tolerance-related functions, such as heat shock proteins and flavonoid metabolic pathways.

Discussion

Maize is an alkali-sensitive crop that is grown on salt-contaminated soil due to soil salinization (Luo et al., 2021). Soil salinization ultimately affects maize yield, but the effect of salinization at the seedling stage is reflected in traits such as seedling length and root fresh weight. In previous studies, the SL, SFW, RFW, SDW, and RDW were used to assess the salinity tolerance of maize, highlighting that these traits showed a significant positive correlation to salinity tolerance (Cui et al., 2014; Luo et al., 2019, 2021). In addition to using the above traits, traits such as RL and RAD were added in this study. A total of 10 relative traits were determined at the seedling stage for GWAS. Among the 200 maize inbred lines that were studied, a large range of phenotypic variation was observed. Except for RAD, which was significantly negatively correlated with other traits, the remaining traits showed significant positive correlations that were consistent with previous studies. The newly added trait, RL, was also

significantly positively correlated with other traits but extremely significantly negatively correlated with RRAD ($r^2 = -0.49$, $p < 0.001$). The mean RRAD of 200 inbred lines was 1.0847, and the average RRL of 200 inbred lines was 0.6664. This indicated that maize may resist external alkali stress by thickening and shortening its roots.

GWAS is now commonly used to detect complex trait variants and predict candidate genes. However, there are still problems with false positives using this method (Wang et al., 2019). RNA-seq has emerged as an important tool for analyzing genome-wide expression patterns (Tai et al., 2016). Nevertheless, potential candidate genes are difficult to identify from the large number of DEGs obtained by RNA-seq. In recent years, GWAS combined in combination with RNA-seq has been used to detect novel genes associated with complex traits in crops. For example, five candidate genes potentially involved in maize root development were predicted using GWAS in combination with RNA-seq. Furthermore, ten candidate genes related to low-temperature tolerance were also identified (Wang et al., 2019; Zhang et al., 2020). This study identified five candidate genes (*Zm00001d038250*, *Zm00001d027619*, *Zm00001d032973*, *Zm00001d001820*, and *Zm00001d001960*) associated with alkali tolerance through this method. Among the identified genes, *Zm00001d038250* and *Zm00001d001960* had

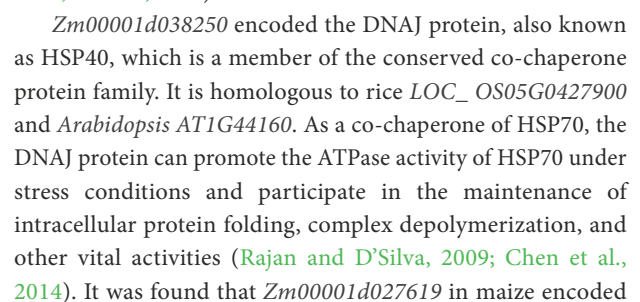


TABLE 4 Candidate DEGs that are consistent between the K4D_ATvsK4D_CK and K3D_ATvsK3D_CK comparison groups.

Gene ID	K4D_ATvsK4D_CK	K3D_ATvsK3D_CK	Gene function	Arabidopsis best hit	Rice best hit
Zm00001d038250	no	up	HSP40/DNAJ peptide-binding protein	DNAJ heat shock family protein	DNAJ protein homolog 1
Zm00001d027619	up	up	beta-amylase	Glycoside hydrolase family 14	beta-amylase 2
Zm00001d032973	up	up	glycerol-3-phosphate acyltransferase7	glycerol-3-phosphate acyltransferase 6	glycerol-3-phosphate 2-O-acyltransferase 6
Zm00001d001820	up	up	protochlorophyllide reductase 1	protochlorophyllide oxidoreductase A	protochlorophyllide reductase A
Zm00001d001960	up	up	Flavanone 3-dioxygenase 1	flavanone 3-hydroxylase	flavanone 3-dioxygenase 1-like

TABLE 5 GO analysis of candidate genes related to the RNA-seq result.

Gene ID	GO
Zm00001d038250	GO:0044444; GO:0008150; GO:0005737; GO:0005575; GO:0044464; GO:0005623; GO:0005829; GO:0003674; GO:0044424; GO:0009987; GO:0005622; GO:0005488; GO:0005515
Zm00001d027619	GO:0008150; GO:0008152; GO:0016787; GO:0003674; GO:0044238; GO:0071704
Zm00001d032973	GO:0008150; GO:0008152; GO:0003824; GO:0016020; GO:0009058; GO:0005575; GO:0016787; GO:0003674; GO:0016740; GO:0009987
Zm00001d001820	GO:0009536; GO:0015979; GO:0044444; GO:0008150; GO:0008152; GO:0003824; GO:0005737; GO:0009058; GO:0005575; GO:0044464; GO:0043227; GO:0005623; GO:0043231; GO:0043229; GO:0043226; GO:0003674; GO:0044424; GO:0005622; GO:0044237
Zm00001d001960	GO:0009628; GO:0008150; GO:0008152; GO:0003824; GO:0009058; GO:0050896; GO:0003674; GO:0005488

TABLE 6 Distribution of genes and pathways related to the RNA-seq results.

Pathway ID	KEGG term	Gene ID
zma01110	Metabolic pathways	Zm00001d027619 Zm00001d001820 Zm00001d001960
zma00500	Starch and sucrose metabolism	Zm00001d027619
zma00860	Porphyrin and chlorophyll metabolism	Zm00001d001820
zma00941	Flavonoid biosynthesis	Zm00001d001960

β-amylase protein and contains an AmyAc conserved domain. However, little research has been presented on β-amylase and its response to abiotic stress using unstructured carbon composed of soluble sugars and starches (Dickman et al., 2019; Zhen et al., 2020).

Next, Zm00001d032973 encoded a glycerol-3-phosphate acyltransferase protein (GPAT). Three types of GPATs have been found in plant cells, which are located in the plastids (including chloroplasts), endoplasmic reticulum, and mitochondria (Murata and Tasaka, 1997; Zheng et al., 2003; Gidda et al., 2009). Among these, GPATs in plastids/chloroplasts are soluble proteins, which affect plants' cold tolerance by affecting the saturation of chloroplast membrane lipid molecules (Murata and Tasaka, 1997; Yokoi et al., 1998; Zhu et al., 2009;

Sui et al., 2017). GPATs distributed in the endoplasmic reticulum and mitochondria are membrane-bound proteins, which participate in the biosynthesis of triacylglycerol (TAGs) in seeds, thereby affecting the accumulation of seed oil (Zheng et al., 2003). Another gene, Zm00001d001820, encoded a protochlorophyllide reductase1 protein (pcr1). A homologous gene in Arabidopsis thaliana, AT5G54190, was shown to be involved in NYEs/SGRs-mediated chlorophyll degradation for detoxification during seed maturation (Li et al., 2017b). The last gene identified in this study, Zm00001d001960, encoded flavanone 3-hydroxylase 1 protein (fht1). It is homologous to rice LOC_OS04G0662600 and Arabidopsis AT3G51240. In general, flavanone 3-hydroxylase 1 improved stress tolerance by upregulating its expression under water-related stress and increasing the content of flavonoids (Klimek-Chodacka et al., 2016; Hodaei et al., 2018).

The two candidate genes (Zm00001d038250 and Zm00001d001960) are thought to function in alkali tolerance, similar to the HSP40 protein and flavanone 3-hydroxylase 1 activity. Zm00001d038250 encoded the HSP40 protein (sHSP), a heat shock protein (HSP), which is a special protein produced by the body in response to stress. Its structure is highly conserved, and it mainly acts as a molecular chaperone to play biological functions such as folding and transportation (Voellmy and Boellmann, 2007). Maize mitochondrial sHsps improved mitochondrial electron transport during salt stress, mainly by protecting NADH: ubiquinone oxidoreductase activity (Hamilton and Heckathorn, 2001). Overexpression

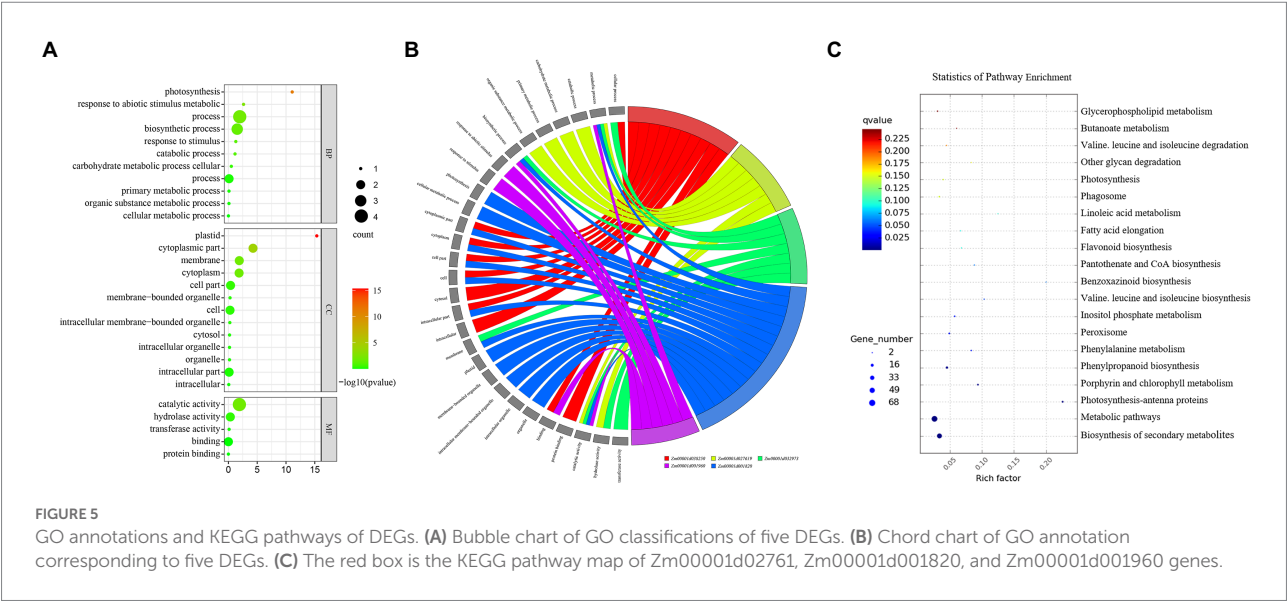


FIGURE 5
GO annotations and KEGG pathways of DEGs. **(A)** Bubble chart of GO classifications of five DEGs. **(B)** Chord chart of GO annotation corresponding to five DEGs. **(C)** The red box is the KEGG pathway map of Zm00001d02761, Zm00001d001820, and Zm00001d001960 genes.

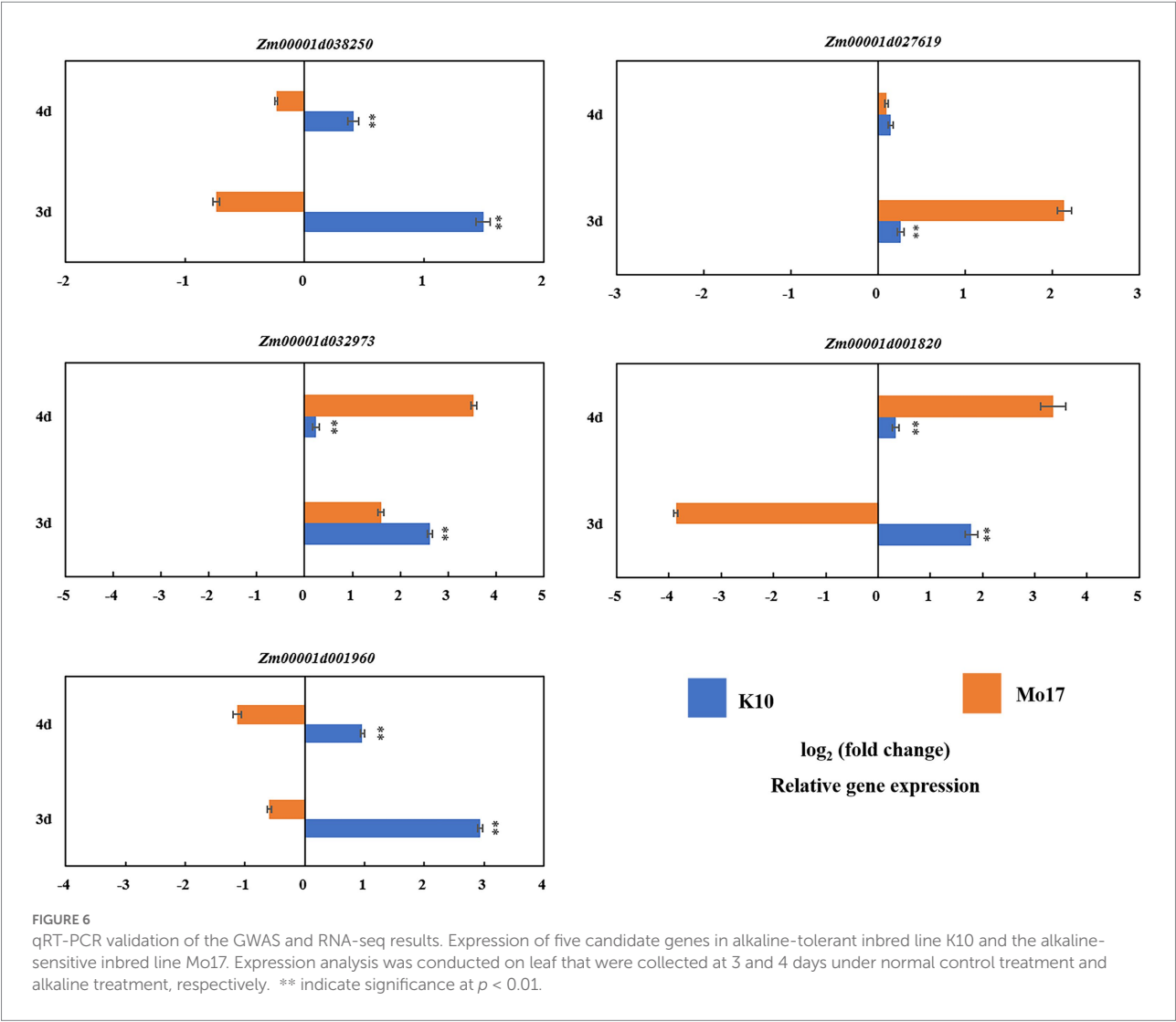


FIGURE 6
qRT-PCR validation of the GWAS and RNA-seq results. Expression of five candidate genes in alkaline-tolerant inbred line K10 and the alkaline-sensitive inbred line Mo17. Expression analysis was conducted on leaf that were collected at 3 and 4 days under normal control treatment and alkaline treatment, respectively. ** indicate significance at $p < 0.01$.

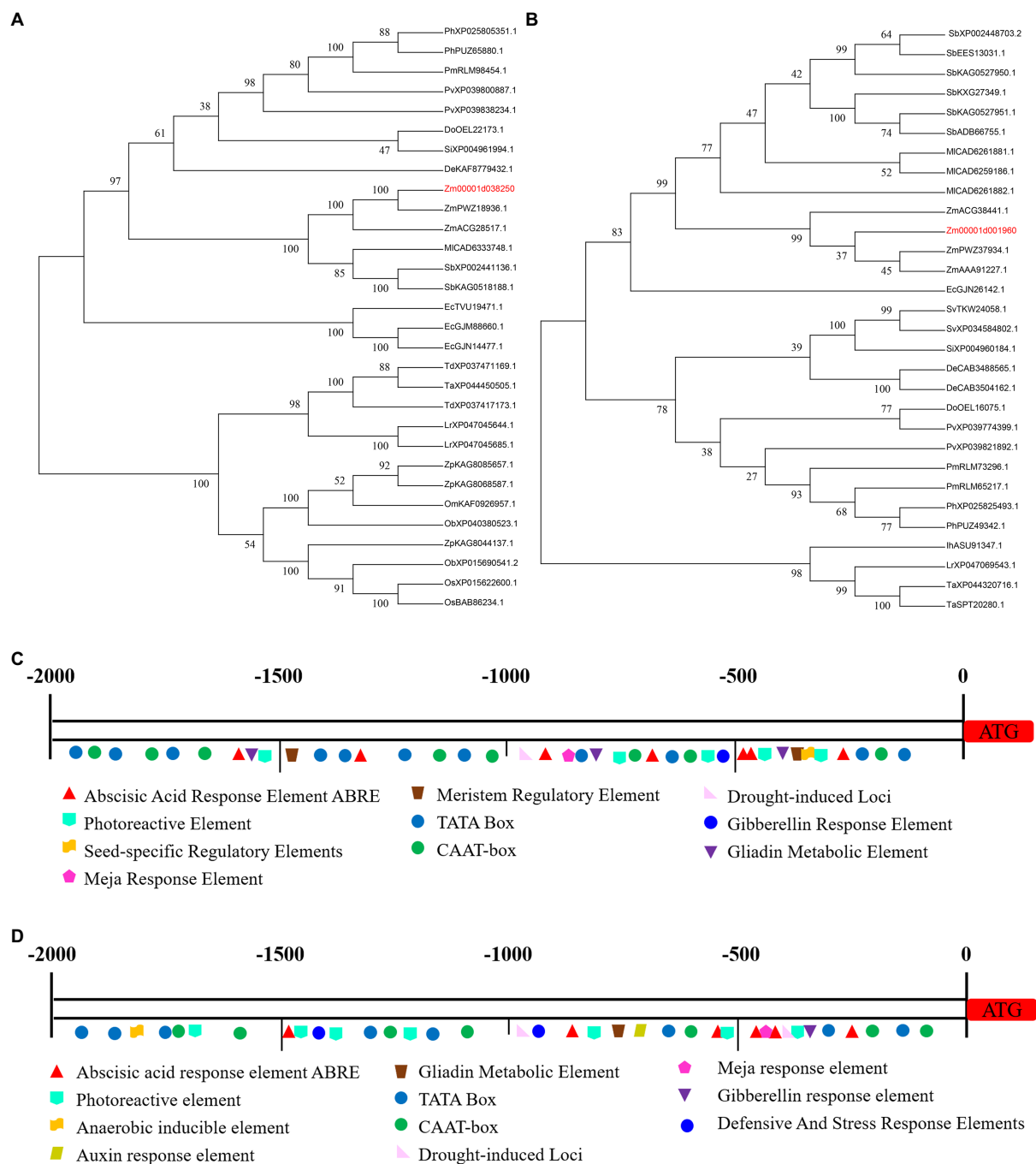


FIGURE 7

The characterization of candidate genes. (A) Evolutionary tree of Zm00001d038250 protein. (B) Evolutionary tree of Zm00001d001960 protein. (C) The promoter region of Zm00001d038250 gene. (D) The promoter region of Zm00001d001960 gene.

of HSP40 in *Arabidopsis* improved salt tolerance by increasing root length under 120 mM NaCl treatment (Zhao et al., 2010). The protein BIL2 in the *Arabidopsis* HSP40 family induced cell elongation during BR signaling by promoting ATP synthesis in mitochondria, thereby increasing inflorescence root length in response to salt and light stress (Bekh-Ochir et al., 2013). Zm00001d038250 and

ZmPWZ18936.1 proteins (NCBI accession number: PWZ18936.1) are evolutionarily on the same branch, and may have some similarities in function. The Zm00001d038250 gene promoter region contained not only CAAT-box core elements but also hormone-, light-, defense-, metabolism-, and stress-related elements (Figures 7A,C; Supplementary Table S6).

Zm00001d001960 putatively encoded flavanone 3-hydroxylase 1 (F3H). Flavonoids are a group of substances with various phenolic structures, which participate in anti-oxidative stress processes, and also participate in some biotic and abiotic stress response processes (Winkler-Shirley, 2002; Panche et al., 2016). Researchers have found that recombinant or overexpression of F3H in different crops could modulate naringenin to alleviate growth inhibition and enhance tolerance to salt stress (Li et al., 2017a; Si et al., 2022). Through proteomic analysis of beet seedlings treated with 50 mM NaCl, 135 differentially expressed proteins, including F3H, were mined (Wu et al., 2018). In addition, it was found that under the combined salt and heat stress in rice, overexpression of F3H increased the kaempferol and quercetin content, and then scavenged reactive oxygen species. The content of heat stress transcription factors (HSFs) and heat shock proteins (HSPs) also increased significantly to improve tolerance (Jan et al., 2021). *Zm00001d001960* is involved in metabolic pathway and flavonoid biosynthesis pathway, which may regulate alkali tolerance through this pathway. The *Zm00001d001960* protein is evolutionarily in the same large clade with the *ZmPWZ37934.1* (NCBI accession number: PWZ37934.1) protein and the *ZmAAA91227.1* (NCBI accession number: AAA91227.1) protein, but its credibility is low, it may be that there is only some similarity in their structure. The *Zm00001d001960* gene promoter region contained not only CAAT-box core elements but also hormone-, light-, and metabolism-related elements. (Figures 7B,D; Supplementary Table S7).

In conclusion, the *Zm00001d038250* and *Zm00001d001960* genes were significantly and differentially expressed in tolerant materials under alkaline conditions based on functional analysis and GO enrichment analysis. It was hypothesized that the HSP40 protein encoded by *Zm00001d038250* could limit the ion transport mode, and the F3H protein encoded by *Zm00001d001960* could improve the antioxidant capacity by metabolizing flavonoids, which, in turn, could affect the alkali tolerance of maize seedlings, though this requires further research. Therefore, these genes are attractive genes for the future study of alkali tolerance in maize seedling stage.

Conclusion

In a population composed of 200 maize inbred lines, we detected 10 alkali tolerance-related traits at the seedling stage and found nine SNPs through GWAS. These SNPs were all related to alkali tolerance during the seedling stage. Through RNA-seq analysis and qPRT-PCR verification, we found that two candidate genes, *Zm00001d038250* and *Zm00001d001960*, were differentially expressed under alkaline treatment. This study provides a genetic basis for molecular-assisted breeding of alkali-tolerant maize seedlings.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ncbi.nlm.nih.gov/>, PRJNA846593.

Author contributions

ZW and HD: conception or design of the work. CL, YJ, and LL: performed the experiment. RZ, MC, and YZ: analyzed the data. CL and YJ: wrote the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This research was supported by the National Natural Science Foundation of China (32072128) and Genetically Modified Organisms Breeding Major Projects of China (2016ZX08003003).

Acknowledgments

We would like to thank Xinhai Li and Jianfeng Weng for providing the plant materials.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary materials

The Supplementary Materials for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.963874/full#supplementary-material>

References

- Abdel Latef, A. A., and Tran, L. S. (2016). Impacts of priming with silicon on the growth and tolerance of maize plants to alkaline stress. *Front. Plant Sci.* 7:243. doi: 10.3389/fpls.2016.00243
- Abreu, I. A., Farinha, A. P., Negrao, S., Goncalves, N., Fonseca, C., Rodrigues, M., et al. (2013). Coping with abiotic stress: proteome changes for crop improvement. *J. Proteomics* 93, 145–168. doi: 10.1016/j.jprot.2013.07.014
- Ambede, J. G., Netondo, G. W., Mwai, G. N., and Musyimi, M. (2012). NaCl salinity affects germination, growth, physiology, and biochemistry of bambara groundnut. *Brazilian J. Plant Physiol.* 24, 151–160. doi: 10.1590/S1677-04202012000300002
- Bekh-Ochir, D., Shimada, S., Yamagami, A., Kanda, S., Ogawa, K., Nakazawa, M., et al. (2013). A novel mitochondrial DnaJ/Hsp40 family protein BIL2 promotes plant growth and resistance against environmental stress in brassinosteroid signaling. *Planta* 237, 1509–1525. doi: 10.1007/s00425-013-1859-3
- 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
- Burton, A. L., Johnson, J. M., Foerster, J. M., Hirsch, C. N., Buell, C. R., Hanlon, M. T., et al. (2014). QTL mapping and phenotypic variation for root architectural traits in maize (*Zea mays* L.). *Theor. Appl. Genet.* 127, 2293–2311. doi: 10.1007/s00122-014-2353-4
- Campbell, S. A., and Nishio, J. N. (2000). Iron deficiency studies of sugar beet using an improved sodium bicarbonate-buffered hydroponic growth system. *J. Plant Nutr.* 23, 741–757. doi: 10.1080/01904160009382056
- Cao, Y. B., Zhang, M., Liang, X. Y., Li, F. R., Shi, Y. L., Yang, X. H., et al. (2020). Natural variation of an EF-hand Ca²⁺-binding-protein coding gene confers saline-alkaline tolerance in maize. *Nat. Commun.* 11, 186–200. doi: 10.1038/s41467-019-14027-y
- Chaparzadeh, N., D'Amico, M. L., Khavari-Nejad, R. A., Izzo, R., and Navari-Izzo, F. (2004). Antioxidative responses of *Calendula officinalis* under salinity conditions. *Plant Physiol. Biochem.* 42, 695–701. doi: 10.1016/j.plaphy.2004.07.001
- Chen, D. H., Huang, Y., Liu, C. L., Ruan, Y., and Shen, W. H. (2014). Functional conservation and divergence of J-domain-containing ZUO1/ZRF orthologs throughout evolution. *Planta* 239, 1159–1173. doi: 10.1007/s00425-014-2058-6
- Chen, S. S., Xing, J. J., and Lan, H. Y. (2012). Comparative effects of neutral salt and alkaline salt stress on seed germination, early seedling growth and physiological response of a halophyte species *Chenopodium glaucum*. *Afr. J. Biotechnol.* 11, 9572–9581. doi: 10.5897/ajb12.320
- Cui, D. Z., Wu, D. D., Somarathna, Y., Xu, C. Y., Li, S., Li, P., et al. (2014). QTL mapping for salt tolerance based on snp markers at the seedling stage in maize (*Zea mays* L.). *Euphytica* 203, 273–283. doi: 10.1007/s10681-014-1250-x
- Dickman, L. T., McDowell, N. G., Grossiord, C., Collins, A. D., Wolfe, B. T., Detto, M., et al. (2019). Homeostatic maintenance of nonstructural carbohydrates during the 2015–2016 El Nino drought across a tropical forest precipitation gradient. *Plant Cell Environ.* 42, 1705–1714. doi: 10.1111/pce.13501
- Dikobe, T. B., Mashile, B., Sinthumule, R. R., and Ruzvidzo, O. (2021). Distinct Morpho-physiological responses of maize to salinity stress. *Am. J. Plant Sci.* 12, 946–959. doi: 10.4236/ajps.2021.126064
- Du, Z., Zhou, X., Ling, Y., Zhang, Z. H., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64–W70. doi: 10.1093/nar/gkq310
- Earl, D. A., and vonHoldt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour.* 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
- Feng, G. X., Zhang, Z. Y., Wan, C. Y., Lu, P. R., and Bakour, A. (2017). Effects of saline water irrigation on soil salinity and yield of summer maize (*Zea mays* L.) in subsurface drainage system. *Agric. Water Manag.* 193, 205–213. doi: 10.1016/j.agwat.2017.07.026
- Gao, Z. W., Han, J. Y., Mu, C. S., Lin, J. X., Li, X. Y., Lin, L. D., et al. (2014). Effects of saline and alkaline stresses on growth and physiological changes in oat (*Avena sativa* L.) seedlings. *Notulae Botanicae Horti Agrobotanici Cluj-Napoca*. 42, 357–362. doi: 10.15835/nbha4229441
- Geng, G., Li, R. R., Stevanato, P., Lv, C. H., Lu, Z. Y., Yu, L. H., et al. (2020). Physiological and transcriptome analysis of sugar beet reveals different mechanisms of response to neutral salt and alkaline salt stresses. *Front. Plant Sci.* 11:571864. doi: 10.3389/fpls.2020.571864
- Gidda, S. K., Shockey, J. M., Rothstein, S. J., Dyer, J. M., and Mullen, R. T. (2009). *Arabidopsis thaliana* GPAT8 and GPAT9 are localized to the ER and possess distinct ER retrieval signals: functional divergence of the dilysine ER retrieval motif in plant cells. *Plant Physiol. Biochem.* 47, 867–879. doi: 10.1016/j.plaphy.2009.05.008
- Guo, J., Li, C. H., Zhang, X. Q., Li, Y. X., Zhang, D. F., Shi, Y. S., et al. (2020). Transcriptome and GWAS analyses reveal candidate gene for seminal root length of maize seedlings under drought stress. *Plant Sci.* 292:110380. doi: 10.1016/j.plantsci.2019.110380
- Hamilton, E. W., and Heckathorn, S. A. (2001). Mitochondrial adaptations to NaCl. Complex I is protected by anti-oxidants and small heat shock proteins, Whereas complex II is protected by Proline and Betaine. *Plant Physiol.* 126, 1266–1274. doi: 10.1104/pp.126.3.1266
- Hartung, W., Leport, L., Ratcliffe, R. G., Sauter, A., Duda, R., and Turner, N. C. (2002). Absciscic acid concentration, root pH and anatomy do not explain growth differences of chickpea (*Cicer arietinum* L.) and lupin (*Lupinus angustifolius* L.) on acid and alkaline soils. *Plant and Soil* 240, 191–199. doi: 10.1023/A:1015831610452
- Hodaei, M., Rahimmalek, M., Arzani, A., and Talebi, M. (2018). The effect of water stress on phytochemical accumulation, bioactive compounds and expression of key genes involved in flavonoid biosynthesis in *Chrysanthemum morifolium* L. *Ind. Crop Prod.* 120, 295–304. doi: 10.1016/j.indcrop.2018.04.073
- Jan, R., Kim, N., Lee, S. H., Khan, M. A., Asaf, S., Asif, S., et al. (2021). Enhanced flavonoid accumulation reduces combined salt and heat stress Through regulation of transcriptional and hormonal mechanisms. *Front. Plant Sci.* 12:796956. doi: 10.3389/fpls.2021.796956
- Jia, T. J., Wang, L. F., Li, J. J., Ma, J., Cao, Y. Y., Lübberstedt, T., et al. (2020). Integrating a genome-wide association study with transcriptomic analysis to detect genes controlling grain drying rate in maize (*Zea mays* L.). *Theor. Appl. Genet.* 133, 623–634. doi: 10.1007/s00122-019-03492-0
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882
- Klimek-Chodacka, M., Sadłowska, K., Kamińska, I., and Barański, R. (2016). Expression of the flavonoid pathway genes in carrot plants tolerant to salt stress. *N. Biotechnol.* 33:S166. doi: 10.1016/j.nbt.2016.06.1294
- Li, P. C., Chen, F. J., Cai, H. G., Liu, J. C., Pan, Q. C., Liu, Z. G., et al. (2015). A genetic relationship between nitrogen use efficiency and seedling root traits in maize as revealed by QTL analysis. *J. Exp. Bot.* 66, 3175–3188. doi: 10.1093/jxb/erv127
- Li, C. C., Liu, S. H., Yao, X. H., Wang, J., Wang, T. L., Zhang, Z. H., et al. (2017a). PnF3H, a flavanone 3-hydroxylase from the Antarctic moss *Pohlia nutans*, confers tolerance to salt stress and ABA treatment in transgenic *Arabidopsis*. *Plant Growth Regul.* 83, 489–500. doi: 10.1007/s10725-017-0314-z
- Li, H., Peng, Z. Y., Yang, X. H., Wang, W. D., Fu, J. J., Wang, J. H., et al. (2013). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484
- Li, X. H., Wang, G. H., Fu, J. J., Li, L., Jia, G. Y., Ren, L. S., et al. (2018). QTL mapping in three connected populations reveals a set of consensus genomic regions for low temperature germination ability in *Zea mays* L. *Front. Plant Sci.* 9:65. doi: 10.3389/fpls.2018.00065
- Li, F. J., Wen, W. E., Liu, J. D., Zhang, Y., Cao, S. H., He, Z. H., et al. (2019). Genetic architecture of grain yield in bread wheat based on genome-wide association studies. *BMC Plant Biol.* 19:168. doi: 10.1186/s12870-019-1781-3
- Li, Z. P., Wu, S. X., Chen, J. Y., Wang, X. Y., Gao, J., Ren, G. D., et al. (2017b). NYEs/SGRs-mediated chlorophyll degradation is critical for detoxification during seed maturation in *Arabidopsis*. *Plant J.* 92, 650–661. doi: 10.1111/tpj.13710
- Li, X. P., Zhou, Z. J., Ding, J. Q., Wu, Y. B., Zhou, B., Wang, R. X., et al. (2016). Combined linkage and association mapping reveals QTL and candidate genes for plant and ear height in maize. *Front. Plant Sci.* 7:833. doi: 10.3389/fpls.2016.00833
- Liang, J. L., Qu, Y. P., Yang, C. G., Ma, X. D., Cao, G. L., Zhao, Z. W., et al. (2014). Identification of QTLs associated with salt or alkaline tolerance at the seedling stage in rice under salt or alkaline stress. *Euphytica* 201, 441–452. doi: 10.1007/s10681-014-1236-8
- Liu, C. L., Hao, Z. F., Zhang, D. G., Xie, C. X., Li, M. S., Zhang, X. C., et al. (2015). Genetic properties of 240 maize inbred lines and identity-by-descent segments revealed by high-density SNP markers. *Mol. Breed.* 35, 146–158. doi: 10.1007/s11032-015-0344-z
- 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
- Luo, M. J., Zhang, Y. X., Chen, K., Kong, M. S., Song, W., Lu, B. S., et al. (2019). Mapping of quantitative trait loci for seedling salt tolerance in maize. *Mol. Breed.* 39, 64–76. doi: 10.1007/s11032-019-0974-7

- Luo, M. J., Zhang, Y. X., Li, J. N., Zhang, P. P., Chen, K., Song, W., et al. (2021). Molecular dissection of maize seedling salt tolerance using a genome-wide association analysis method. *Plant Biotechnol. J.* 19, 1937–1951. doi: 10.1111/pbi.13607
- Ma, L. L., Qing, C. Y., Frei, U., Shen, Y. O., and Lübberstedt, T. (2020). Association mapping for root system architecture traits under two nitrogen conditions in germplasm enhancement of maize doubled haploid lines. *Crop J.* 8, 213–226. doi: 10.1016/j.cj.2019.11.004
- Miculan, M., Nelissen, H., Ben Hassen, M., Marroni, F., Inzè, D., EnricoPè, M., et al. (2021). A forward genetics approach integrating genome-wide association study and expression quantitative trait locus mapping to dissect leaf development in maize (*Zea mays*). *Plant J.* 107, 1056–1071. doi: 10.1111/tj.15364
- Mishra, P., Bhoomika, K., and Dubey, R. S. (2013). Differential responses of antioxidative defense system to prolonged salinity stress in salt-tolerant and salt-sensitive Indica rice (*Oryza sativa* L.) seedlings. *Protoplasma* 250, 3–19. doi: 10.1007/s00709-011-0365-3
- Murata, N., and Tasaka, Y. (1997). Glycerol-3-phosphate acyltransferase in plants. *Biochim. Biophys. Acta* 1348, 10–16. doi: 10.1016/s0005-2760(97)00115-x
- Panche, A. N., Diwan, A. D., and Chandra, S. R. (2016). Flavonoids: an overview. *J. Nutr. Sci.* 5:e47. doi: 10.1017/jns.2016.41
- Pardo, J. M., and Quintero, F. J. (2002). Plants and sodium ions: keeping company with the enemy. *Genome Biol.* 3:reviews1017. doi: 10.1186/gb-2002-3-6-reviews1017
- Peng, Y. L., Gao, Z. W., Gao, Y., Liu, G. F., Sheng, L. X., and Wang, D. L. (2008). Eco-physiological characteristics of alfalfa seedlings in response to various mixed salt-alkaline stresses. *J. Integr. Plant Biol.* 50, 29–39. doi: 10.1111/j.1744-7909.2007.00607.x
- Rajan, V. B., and D'Silva, P. (2009). *Arabidopsis thaliana* J-class heat shock proteins: cellular stress sensors. *Funct. Integr. Genomics* 9, 433–446. doi: 10.1007/s10142-009-0132-0
- Ruanjaichon, V., Khammona, K., Thunnon, B., Suriharn, K., Kerdri, C., Aesomnuk, W., et al. (2021). Identification of gene associated with sweetness in corn (*Zea mays* L.) by genome-wide association study (GWAS) and development of a functional SNP marker for predicting sweet corn. *Plants* 10, 1239–1250. doi: 10.3390/plants10061239
- Sajjad, M., Ma, X. L., Habibullah Khan, S., Shoaib, M., Song, Y. H., Yang, W. L., et al. (2017). *TaFlo2-A1*, an ortholog of rice *Flo2*, is associated with thousand grain weight in bread wheat (*Triticum aestivum* L.). *BMC Plant Biol.* 17:164. doi: 10.1186/s12870-017-1114-3
- Schachtman, D., and Liu, W. (1999). Molecular pieces to the puzzle of the interaction between potassium and sodium uptake in plants. *Trends Plant Sci.* 4, 281–287. doi: 10.1016/S1360-1385(99)01428-4
- Shi, D. C., and Sheng, Y. M. (2005). Effect of various salt-alkaline mixed stress conditions on sunflower seedlings and analysis of their stress factors. *Environ. Exp. Bot.* 54, 8–21. doi: 10.1016/j.envexpbot.2004.05.003
- Si, C., Dong, W., da Silva, J. A. T., He, C. M., Yu, Z. M., Zhang, M. Z., et al. (2022). Functional analysis of Flavanone 3-hydroxylase (F3H) from *Dendrobium officinale*, which confers abiotic stress tolerance. *Hortic. Plant J.* doi: 10.1016/j.hpj.2022.03.006 (in press).
- Sui, N., Tian, S. S., Wang, W. Q., Wang, M. J., and Fan, H. (2017). Overexpression of Glycerol-3-phosphate Acyltransferase from *Suaeda salsa* improves salt tolerance in *Arabidopsis*. *Front. Plant Sci.* 8:1337. doi: 10.3389/fpls.2017.01337
- Sun, J., Xie, D. W., Zhang, E. Y., Zheng, H. L., Wang, J. G., Liu, H. L., et al. (2019). QTL mapping of photosynthetic-related traits in rice under salt and alkali stresses. *Euphytica* 215, 147–161. doi: 10.1007/s10681-019-2470-x
- Tai, H. H., Lu, X., Opitz, N., Marcon, C., Paschold, A., Lithio, A., et al. (2016). Transcriptomic and anatomical complexity of primary, seminal, and crown roots highlight root type-specific functional diversity in maize (*Zea mays* L.). *J. Exp. Bot.* 67, 1123–1135. doi: 10.1093/jxb/erv513
- Voellmy, R., and Boellmann, F. (2007). Chaperone regulation of the heat shock protein response. *Adv. Exp. Med. Biol.* 594, 89–99. doi: 10.1007/978-0-387-39975-1_9
- Wang, H. M., Wei, J., Li, P. C., Wang, Y. Y., Ge, Z. Z., Qian, J. Y., et al. (2019). Integrating GWAS and gene expression analysis identifies candidate genes for root morphology traits in maize at the seedling stage. *Genes* 10, 773–789. doi: 10.3390/genes10100773
- Weng, J. F., Xie, C. X., Hao, Z. F., Wang, J. J., Liu, C. L., Li, M. S., et al. (2011). Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. *PLoS One* 6:e29229. doi: 10.1371/journal.pone.0029229
- Winkler-Shirley, B. (2002). Biosynthesis of flavonoids and effects of stress. *Curr. Opin. Plant Biol.* 5, 218–223. doi: 10.1016/S1369-5266(02)00256-X
- Wu, G. Q., Wang, J. L., Feng, R. J., Li, S. J., and Wang, C. M. (2018). iTRAQ-based comparative proteomic analysis provides insights into molecular mechanisms of salt tolerance in sugar beet (*Beta vulgaris* L.). *Int. J. Mol. Sci.* 19, 3866–3886. doi: 10.3390/ijms19123866
- Xu, J. J., Liu, T., Yang, S. C., Jin, X. Q., Qu, F., Huang, N., et al. (2019). Polyamines are involved in GABA-regulated salinity-alkalinity stress tolerance in muskmelon. *Environ. Exp. Bot.* 164, 181–189. doi: 10.1016/j.envexpbot.2019.05.011
- Yan, J. B., Warburton, M., and Crouch, J. (2011). Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. *Crop. Sci.* 51, 433–449. doi: 10.2135/cropsci2010.04.0233
- Yang, Y. Q., Wu, Y. J., Ma, L., Yang, Z. J., Dong, Q. Y., Li, Q. P., et al. (2019). The Ca²⁺ sensor SCaBP3/CBL7 modulates plasma membrane H⁺-ATPase activity and promotes alkali tolerance in *Arabidopsis*. *Plant Cell* 31, 1367–1384. doi: 10.1105/tpc.18.00568
- Yokoi, S., Higashi, S., Kishitani, S., Murata, N., and Toriyama, K. (1998). Introduction of the cDNA for *Arabidopsis* glycerol-3-phosphate acyltransferase (GPAT) confers unsaturation of fatty acids and chilling tolerance of photosynthesis on rice. *Mol. Breed.* 4, 269–275. doi: 10.1023/A:1009671231614
- Zhang, C. X., Jin, F. X., Li, S. F., Liu, W. P., Ma, X. J., Yang, S., et al. (2018). Fine mapping of major QTLs for alkaline tolerance at the seedling stage in maize (*Zea mays* L.) through genetic linkage analysis combined with high-throughput DNA sequencing. *Euphytica* 214, 120–132. doi: 10.1007/s10681-018-2190-7
- Zhang, H., Zhang, J. Y., Xu, Q. Y., Wang, D. D., Di, H., Huang, J., et al. (2020). Identification of candidate tolerance genes to low-temperature during maize germination by GWAS and RNA-seq approaches. *BMC Plant Biol.* 20:333. doi: 10.1186/s12870-020-02543-9
- Zhang, C. S., Zhou, Z. Q., Yong, H. J., Zhang, X. C., Hao, Z. F., Zhang, F. J., et al. (2017). Analysis of the genetic architecture of maize ear and grain morphological traits by combined linkage and association mapping. *Theor. Appl. Genet.* 130, 1011–1029. doi: 10.1007/s00122-017-2867-7
- Zhao, Z., Zhang, W., Yan, J., Zhang, J., Liu, Z. L. X., and Yi, Y. (2010). Over-expression of *Arabidopsis* DnaJ (Hsp40) contributes to NaCl-stress tolerance. *Afr. J. Biotechnol.* 9, 972–978. doi: 10.5897/ajb09.1450
- Zhen, F. X., Zhou, J. J., Mahmood, A., Wang, W., Chang, X. N., Liu, B., et al. (2020). Quantifying the effects of short-term heat stress at booting stage on nonstructural carbohydrates remobilization in rice. *Crop J.* 8, 194–212. doi: 10.1016/j.cj.2019.07.002
- Zheng, Z. F., Xia, Q., Dauk, M., Shen, W. Y., Selvaraj, G., and Zou, J. T. (2003). *Arabidopsis AtGPAT1*, a member of the membrane-bound glycerol-3-phosphate acyltransferase gene family, is essential for tapetum differentiation and male fertility. *Plant Cell* 15, 1872–1887. doi: 10.1105/tpc.012427
- Zhu, J. K. (2016). Abiotic stress signaling and responses in plants. *Cell* 167, 313–324. doi: 10.1016/j.cell.2016.08.029
- Zhu, S. Q., Zhao, H., Zhou, R., Ji, B. H., and Dan, X. Y. (2009). Substrate selectivity of glycerol-3-phosphate acyl transferase in rice. *J. Integr. Plant Biol.* 51, 1040–1049. doi: 10.1111/j.1744-7909.2009.00876.x



OPEN ACCESS

EDITED BY

Xia Xin,
Institute of Crop Sciences (CAAS),
China

REVIEWED BY

Jianfeng Weng,
Institute of Crop Sciences (CAAS),
China
Yan He,
China Agricultural University, China

*CORRESPONDENCE

Zhenhua Wang
zhenhuawang_2006@163.com
Hong Di
dihongdh@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION

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

RECEIVED 08 June 2022

ACCEPTED 28 June 2022

PUBLISHED 19 July 2022

CITATION

Song Y, Li C, Zhu Y, Guo P, Wang Q,
Zhang L, Wang Z and Di H (2022)
Overexpression of *ZmIPT2* gene delays
leaf senescence and improves grain
yield in maize.
Front. Plant Sci. 13:963873.
doi: 10.3389/fpls.2022.963873

COPYRIGHT

© 2022 Song, Li, Zhu, Guo, Wang,
Zhang, Wang and Di. 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(s) 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 *ZmIPT2* gene delays leaf senescence and improves grain yield in maize

Yongfeng Song[†], Chunxiang Li[†], Yong Zhu, Pei Guo, Qi Wang,
Lin Zhang, Zhenhua Wang* and Hong Di*

Key Laboratory of Germplasm Enhancement, Physiology and Ecology of Food Crops in Cold
Region, Ministry of Education, Northeast Agricultural University, Harbin, China

Cytokinins (CTKs) are a major phytohormone group that are significant in the promotion of cellular division, growth, and divergence. *Isopentenyl transferase (IPT)* regulates a rate-limiting step in plant CTK synthesis, promotes the synthesis of isopentenyl adenonucleotides from 5-AMP and isopentenyl pyrophosphate, and then converts both these chemicals into various CTKs. Here, the full-length cDNA of *ZmIPT2*, which encodes 322 amino acids, was isolated and was introduced into a maize inbred line by *Agrobacterium*-mediated transformation. In both controlled environments and field experiments, the overexpression of *ZmIPT2* gene in the transformed plants delayed leaf senescence. Compared to the receptor line, the transgenic maize lines retained higher chlorophyll levels, photosynthetic rates, and cytokinin content for an extended period of time, and produced significantly higher grain yield by a margin of 17.71–20.29% under normal field planting conditions. Subsequently, ten possible genes that interacted with *ZmIPT2* were analyzed by qRT-PCR, showing that the expression pattern of *GRMZM2G022904* was consistent with *ZmIPT2* expression. Through comprehensive analysis, we screened for transgenic lines with stable inheritance of *ZmIPT2* gene, clear functional efficiency, and significant yield improvement, in order to provide theoretical basis and material support for the breeding of new high-yield transgenic maize varieties.

KEYWORDS

maize, leaf senescence, *ZmIPT2* gene, grain yield, overexpression

Introduction

Senescence is a degradative process of plant tissue structures (Rivero et al., 2007). To improve overall plant health, leaves should be engineered for delayed senescence to permit the capture of sunlight energy for a longer duration, allowing for more efficient transformation of light into photosynthates (Sykorova et al., 2008). Additionally, a delay in senescence would also allow tissues to degenerate more slowly. In doing so, the stored resources can be systematically released to the appropriate sink tissues. Advantages of

deferred senescence involve a better maintenance of photosynthetic rate, raised plant biomass, improved drought tolerance, and higher seed yield (McCabe et al., 2001; Chang et al., 2003; Rivero et al., 2010).

Cytokinins (CTKs) were thought to have evolved to coordinate the endogenous developmental processes and adaptive responses in plants (Li et al., 2016). Several links have been established between CTK level, pool strength, and source-pool conversion by regulating photoassimilate distribution (Brenner and Cheikh, 1995; Roitsch and Rainer, 2000; Kuiper, 1993). It is generally believed that the CTK-mediated delay of leaf senescence was related to sink/source regulation, and that part of this relationship is mediated by the cell wall invertase (CWINV) protein (Roitsch and Rainer, 2000). Scientists found that leaf senescence was delayed and expression level of CWINV was increased in *PSAG12:IPT* transgenic tobacco, indicating that CWINV was closely related to senescence. The CWINV inhibitor protein was expressed under the CTK inducible promoter, and there was no delay in leaf senescence after CTK treatment (Balibrea Lara et al., 2004), indicating that the increased presence of CWINV was a necessary condition for CTK to delay senescence. This, at least, partially explained the molecular mechanisms of CTK in delaying leaf senescence (Peng et al., 2021).

The function of CTKs in deferring leaf senescence has been reported for several plant species (Gan and Amasino, 1995; Robson et al., 2004; Rivero et al., 2007; Sykorova et al., 2008; Peleg et al., 2011; Kant et al., 2015). Gan and Amasino were the first to apply the senescence-specific promoter, *SAG12*, fused with the *IPT* gene to explore the function of CTKs in deferring leaf senescence. The transgenic tobacco plants that were created in these studies reported higher biomass and yield (Gan and Amasino, 1995). Further studies on transgenic tobacco found that the expression of the phosphoenolpyruvate gene was regulated by *P_{SARK}*, a gene specifically induced under plant aging and stress conditions (Delatorre et al., 2012). Regulation of the phosphoenolpyruvate gene could maintain a high concentration of cytokinin in plants under water stress, and improve the tolerance of plants to water stress and prolong the photosynthetic life of plant leaves (Rivero et al., 2007). Decreased local CTK production in roots or leaves, which are typically plentiful sources of CTKs, were assumed to initiate leaf senescence (Singh, 1992; Hirose et al., 2008; Kudo et al., 2010). Isopentenyladenine (iP), and one of its glucosides, iP9G, were both capable of delaying chlorophyll degradation in detached cotyledons senescence assays (Hallmark and Rashotte, 2020).

From an agronomic point of view, several studies link increased CTKs levels to seed yield in rice, soybean, and maize. In some varieties of rice, the cytokinin oxidase gene (*OsCTKx2*) was associated with the quantitative trait locus (QTL) for grain production (Ashikari et al., 2005). In soybean,

CTKs have been shown to play a significant function on pod growth and flowering (Huff and Dybing, 1980; Ghiasi et al., 1987; Wiebold and Panciera, 1990; Westgate and Peterson, 1993; Nagel, 2001). In maize, CTKs' exogenous application has been similarly revealed to increase the yield stability of heat stressed plants (Cheikh and Jones, 1994). Therefore, CTKs are involved in a large amount of quantitative and qualitative components of yield. Their participation in a huge quantity of regulatory roles in plants give much incentive to unravel biosynthesis of CTKs and signal pathways (Kieber and Schaller, 2014).

The production of isopentenyl transferase (*IPT*), a key enzyme of CTK biosynthesis, catalyzes the first step of *de novo* synthesis of cytokinin, promotes the synthesis of isopentenyl adenonucleotides from 5-AMP and isopentenyl pyrophosphate, and then converts both these chemicals into various CTKs. The *IPT* gene was first identified in *Agrobacterium tumefaciens* (Gan and Amasino, 1995). Since the discovery of the *IPT* gene, there have been a large number of reports using transgenic technology to study the use of this gene to improve yield, anti-aging, and disease resistance (Noh and Amasino, 1999). The mechanism of inducing the *IPT* gene expression to produce endogenous CTK to delay leaf senescence may be related to promoting the accumulation of stress response related proteins and antioxidant enzymes (Peng et al., 2021). Studies have shown that the heterologous expression of *IPT* under the transcription control of senescence-associated receptor-like kinase (SARK) promoter in wheat can improve plant drought tolerance by delaying cell senescence. Overexpressing *IPT* showed a water tolerance phenotype in different environments: transgenic TR1 and TR4 wheat showed delayed senescence and increased yield under sufficient water conditions (Bez nec et al., 2021), similar to the reported effect of *IPT* expression in rice, tobacco, and maize (Rivero et al., 2007; Peleg et al., 2011; Qin et al., 2011; Decima Oneto et al., 2016). Interestingly, in transgenic cauliflower containing the *IPT* gene, a tetrapeptide-like repeat was found, which could induce protein folding, SOD and stress response. The protein folding and carbon fixation related protein levels of this transgenic cauliflower were higher than those of non-transgenic cauliflower, and the iron SOD and APX activities were also significantly higher than those of non-transgenic cauliflower (Liu et al., 2011). The transgenic plants had an increased number of stress-related proteins and molecular chaperones, induced the expression of stress-related genes, and protected cells during aging in contemporary plant breeding programs (Peng et al., 2021).

These techniques are considered new strategies for engineering CTKs for better crop production in agricultural systems (Jameson and Song, 2016; Kieber and Schaller, 2018). In addition, *Arabidopsis* researchers have shown that the manipulation of CTK levels via *IPTs* can enhance salt stress tolerance and drought (Li et al., 2016), laying the groundwork

for similar studies in agriculturally significant plants. In addition, a growing number of studies have shown that during stress, the expression of the *IPT* gene can induce plants to produce various resistant or active substances, which can produce CTK responses, coordinate, and interact with other hormones. The production of various resistance and active substances in plants can enhance photosynthesis. Various metabolites can be synthesized to promote the response of plant cells to various adverse environments or diseases and insect pests. Moreover, agronomic traits have been significantly improved by the induction of the *ITP* gene (Li and Zhao, 2011), which provided a novel base for future molecular breeding of crop plants (Jameson and Song, 2016).

In crops such as lettuce (McCabe et al., 2001), rice (Jin et al., 2002), wheat (Sykorova et al., 2008), and ramie (An et al., 2017), CTK manipulation not only delays the aging of leaves, but also improves agronomic traits and stress resistance of crops. Gan and Amasino transferred the chimeric genes *SAG12* and *IPT* into tobacco and found that the transgenic plants showed higher biomass and yield (Gan and Amasino, 1995). In rice, *IPT*-induced CTKs synthesis reduced environmental stress consequences on photosynthesis and yield and maintained nitrogen (N) acquisition (Reguera et al., 2013). Overexpression of *OsIPT9* in rice can increase the CTK level of caryopsis development, resulting in enhanced grain filling of large and multi-panicle rice varieties, thus increasing yield (Panda et al., 2018). Transgenic wheat containing the *IPT* gene driven by *AtMYB2xs* promoter had improved yield under both sufficient water and water stress conditions (Joshi et al., 2019). Regulating *IPT* by using the *AtMYB32* promoter can improve rape yield under both drought and conventional irrigation conditions (Kant et al., 2015). Multiple species showed improved yield by increasing *IPT* expression levels under drought conditions, indicating that if the expression of *ITP* can be fully controlled, it could be a key driver of yield in crop systems.

In maize, eight *IPT* genes have been found in the genome, all of which have complete open reading frames (ORFs) except *ZmIPT3*. Among the completed ORFs, the *ZmIPT2* gene encodes a 322 amino acid protein and is stably expressed at the transcriptional level, in addition to being the highest expressed protein (Takei et al., 2001). Some studies have proved that in maize tissue, the content of CTKs was directly proportional to the expression level of *ZmIPT2*. When the expression level of *ZmIPT2* increased, the CTKs content also increased. Therefore, the expression of this gene was considered to be related to the content of cytokinin. The *ZmIPT2* gene played a very important role in overall seed development, from endosperm cell division to embryo development, and played a role in germplasm strength. It can also delay leaf senescence and improve yield (Brugiere et al., 2008).

The endeavor to comprehend the molecular basis of this increase in yield owing to deferred senescence and the

possible involvement of CTKs in this process will be useful in aiding breeders and scientists alike. Our laboratory cloned *ZmIPT2* gene from early maize, and transformed the maize using *Agrobacterium*, showing that the transgenic offspring held their leaves during the regular leaf senescence stage, and maintained a brighter green color in their leaves. The transgenic offspring also had an increased yield, although the genetic instability phenomenon was observed. Therefore, we reconstructed the monocotyledonous plant expression vector of this gene, and transformed maize inbred lines by *Agrobacterium tumefaciens* infection of young embryos to obtain T2 transgenic lines. We clarified the basic characteristics of this gene through bioinformatic analysis, examined the expression pattern under different growth stages, and analyzed the physiological indexes and yield traits of transgenic lines. The interacting genes of *ZmIPT2* were then screened by qRT-PCR. The transgenic maize line that we have generated may be of great value as a genetic resource for further high-yield maize breeding.

Materials and methods

Plant materials

The maize inbred line Zheng 58 was used to clone the target gene and the maize inbred line C01 was used as the receptor material for genetic transformation. The material for Zheng 58 was obtained from the Maize Research Institute of Northeast Agricultural University. The inbred line C01 used as a transgenic receptor was provided by the Life Science and Technology Center of China Seed Group Co., Ltd. According to the relevant laws of China, after the experiment the seeds were preserved on-site and the plants were properly disposed of.

Agrobacterium tumefaciens strain LBA4404 came from our laboratory. *Escherichia coli* strain DH5a came from Beijing Quanshijin company. The basic plant expression vector used in the transformation experiments was pEAST-T1, with the ubiquitin promotor, T-NOS terminator, and the CaMV35S promoter driven bar gene selection marker, which was provided by the Institute of Crop Research, Chinese Academy of Agricultural Sciences. Taq DNA polymerase was purchased from Beijing Quanshijin biological Co., Ltd.

Methods

The characterization and cloning of the *ZmIPT2* gene

Bioinformatics analysis of *ZmIPT2* gene

The *ZmIPT2* gene (Gene ID: GRMZM2G084462) was localized in bin 2.04 region based on B73 RefGen_v2 genome-wide data. Analysis of the *ZmIPT2* ORF was performed with

ORF finder¹. The second-order structure of the *ZmIPT2* protein was predicted using SOMPA², with SWISS – MODELL³ to predict the protein tertiary structure. The Conserved Domain Database website⁴ was used for protein structure analysis, and Phytozome V 12.1 was used for the other species of IPT family protein sequences. The system evolution tree was constructed by using ClustalX 1.83 and MEGA 5 software, with the bootstrap value adjusted to 1000. The physicochemical properties and hydrophobicity of the *ZmIPT2* protein were analyzed with protparam⁵ and ProtScale⁶. The subcellular localization of *ZmIPT2* protein was predicted using Softberry⁷. The *cis*-acting elements of the *ZmIPT2* gene flanking sequences were identified using the PlantCARE program⁸.

Gene cloning and plant expression vector construction

Using the accession number EU263126.1 on GenBank, the mRNA sequence of *ZmIPT2* gene was downloaded and PCR primers were designed. Total RNA was extracted from leaves of maize inbred line Zheng 58 with the plant RNA Kit (Beijing Tiangen, China) and reverse transcribed to obtain the first strand cDNA. Under the following conditions, the cDNA was amplified by PCR with specific primers of *ZmIPT2* using KOD plus NEO (Toyobo) polymerase mixture. The recovered PCR products were sent to Shanghai Sangong for sequencing.

Total RNA was extracted from leaves of the inbred maize line Zheng 58 with the Plant RNA Kit (Beijing Tiangen, China) following the manufacturer's protocols. Extracted RNA was treated with DNase I (Omega, GA, United States) to remove residual DNA. The first-strand cDNA was synthesized from 1 µg of total RNA using 1 µL (200 U) PrimeScript™ RT Enzyme Mix I (TaKaRa) according to the manufacturer's protocol, to produce the full-length cDNA of *ZmIPT2*. PCR was performed on the cDNA using specific primers for *ZmIPT2* using the KOD-Plus-Neo (TOYOBO) polymerase mix under the following conditions. The amplification product was purified from 1.0% agarose gel and cloned into the pEASY-T1 Simple vector (Transgen) to be sent for sequencing. The primers are listed in **Supplementary Table 1**. The full sequence of the *ZmIPT2* cDNA clone used in this study can be found in GenBank (GenBank Accession No. EU263126.1).

For generating the transgenic lines overexpressing *ZmIPT2*, the coding sequence of *ZmIPT2* was subcloned from the pEAST-T1 vector into the pEC-Ubi-Tnos-Bar vector, which consisted of a maize constitutive ubiquitin promoter, Bar gene with the CaMV 35S promoter, and a Tnos terminator (Sambrook et al., 1982), using to the standard protocol for the In-Fusion Cloning Kit [Takara Biomedical Technology (Beijing) Co., Ltd]. The recombinant plasmid was finally transformed into *Agrobacterium tumefaciens* LBA4404 using the freeze-thaw method (**Supplementary Figure 1**).

Maize transformation

In this study, maize inbred line C01 was used as the starting genetic material. Maize transformation was performed according to the method published by Li et al. (2008). C01 seeds were surface sterilized and cultured on sterile MS agar at 28°C until etiolated seedlings grew to 3.0–4.0 cm. The maize shoot tip with exposed meristem was immersed in the transformed *Agrobacterium* suspension, which was grown a logarithmic stage, and then diluted to an OD600 of 0.8. The buds were soaked at 0.05 MPa for 10 min and then co-cultured on modified MS agar in the dark at 24°C for 3 days. The seedlings were then transplanted into pots and transferred to the greenhouse to continue to grow. Transgenic seedlings sprayed with 0.1% herbicide Basta were detected by PCR, and seeds were harvested from positive plants (T0). T1 plants were grown from harvested seeds and self-pollinated to produce the T2 generation. Because of the vector contains a screening marker gene Bar gene, the positive transgenic plants of each generation had the integration and stability of the *ZmIPT2* gene by verified by PCR and Bar protein test strip. The expression level of *ZmIPT2* and copy number in transgenic plants was confirmed by qRT-PCR.

Molecular detection of transgenic maize lines PCR and RT-PCR

The leaf of DNA was extracted by CTAB method (Murray and Thompson, 1980), and specific primers were designed according to the target gene sequence and terminator sequence. One end of the primer was located inside the target gene, and the other end was located inside the terminator. The primers of *ZmIPT2* gene were F (5'-TTTAGCCCTGCCTTCA-3'), R (5'-AACCCATCTCATAAATAACG-3'), and bar gene specific primers BF1 (5'-CCATCTCAACCACTACATCG-3') and BR1 (5'-AGCTGCCAGAAACCCACGT-3'). PCR reactions were set up in volumes of 20 µL, as recommended by the manufacturer. The thermal cycling conditions were 95°C for 5 min, followed by 30 cycles of 95°C for 30 s, 50°C for 30 s, 72°C for 70 s, and 72°C for 10 min. Total RNA was extracted using Trizol reagent (Tiangen, Beijing, China) from 100 mg of maize seedling leaves. The RNA

1 <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>

2 https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html

3 <https://swissmodel.expasy.org/>

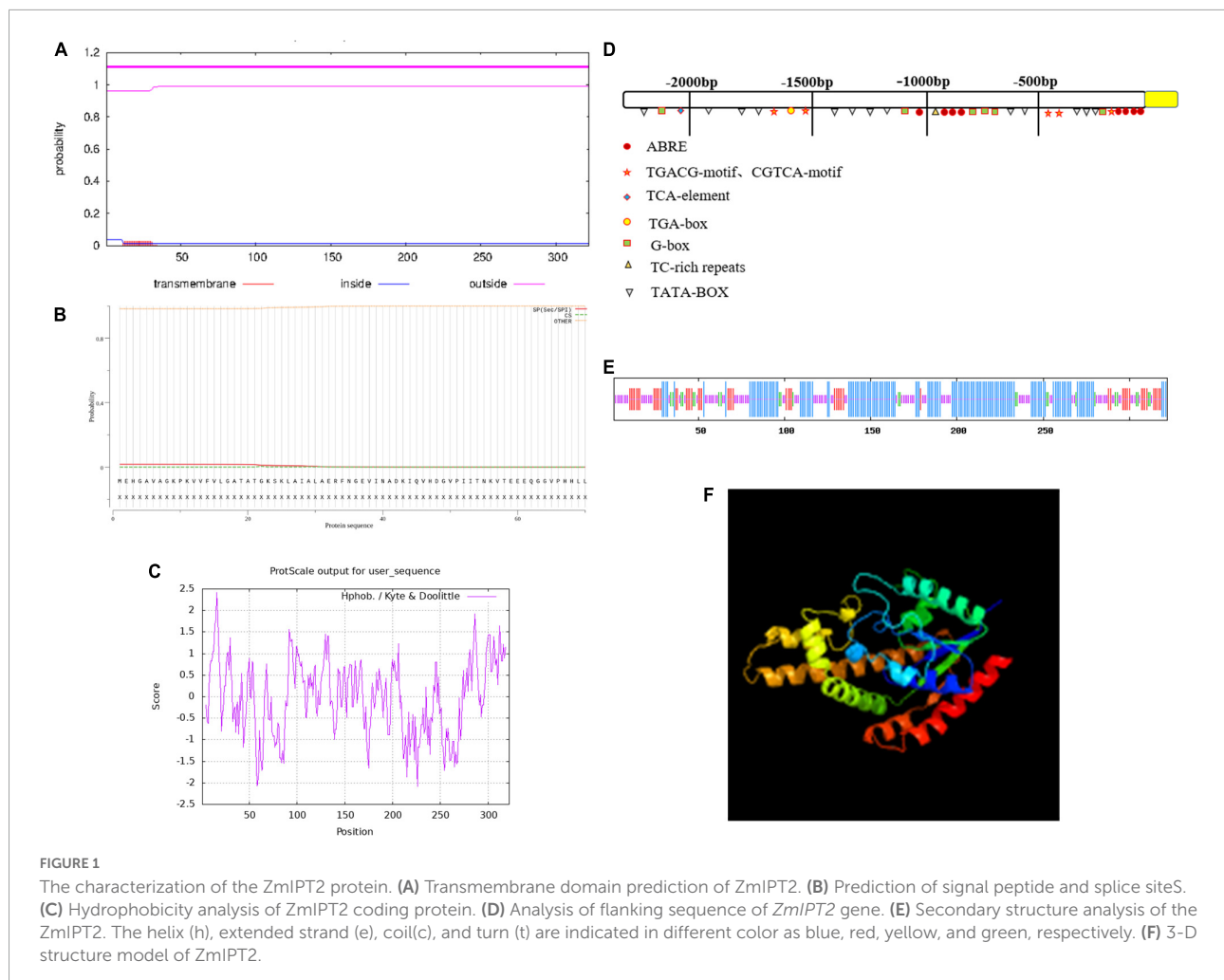
4 <http://www.ncbi.nlm.nih.gov/cdd>

5 <http://www.expasy.org/tools/protparam>

6 <http://web.expasy.org/protscale/>

7 <http://www.softberry.com/>

8 <http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>



was DNase-treated, and 500 ng of this treated RNA was used for inverse transcription with the RT Reagent Kit (Quanshijin, Beijing, China), following the manufacturer's protocols. The generated cDNA samples were diluted fivefold to serve as templates for the subsequent PCR. The primers Actin-F (5'-GTTGGGCGTCCTCGTCA-3') and Actin-R (5'-TGGGTCATCTTCTCCCTGTT-3') were designed for qRT-PCR, and three technical replicates were used for each of the RT-PCR experiments in this research.

Expression analysis by real-time quantitative PCR

The V6 (sixth leaf), R1 (silking), R2 (blister stage), and R6 (physical maturity) leaves of maize were collected at different growth stages. The Trizol reagent (Quanshijin, Beijing, China) was used to extract total RNA. For each sample, the Transscript® One step gDNA Removal Kit mix (Transgen Biotech, Beijing, China) was combined with 500 ng RNA to synthesize cDNA. qRT-PCR was performed using the TransStart Tip Green qPCR SuperMix fluorescent quantitative Kit (Genetically Modified Biotechnology

Company, Beijing, China). The Actin1 and *ZmIPT2* gene-specific primer sequences used for qRT-PCR are located **Supplementary Table 1**. The gene expression levels were calculated using the $2^{-\Delta\Delta Ct}$ formula method (Livak and Schmittgen, 2001). For each sample, three biological replicates were included, with three technical replicates for each. The subsequent analysis was performed by using the obtained qRT-PCR data.

The copy number analysis of transgenic plants

The transgene copy number was detected by qRT-PCR. Actin1 (Gene ID: 100282267), which is a single copy gene in maize, was used as the endogenous reference gene. According to the principle that the CT value obtained by PCR reaction is linearly inversely proportional to the logarithm of the number of initial templates, a quantitative standard curve with a correlation coefficient of more than 0.99 between the internal reference gene actin and the target gene *ZmIPT2* was prepared. After diluting the DNA concentration of each sample to be tested by five times, the qRT-PCR reaction was

conducted to obtain CT values and calculate the number of initial templates of each sample according to the standard curve. Since the actin gene is a single copy in maize, the ratio of the logarithm of the initial template number of *ZmIPT2* gene and actin gene is the copy number of the target gene in maize (Han et al., 2016).

Detection of physiological and biochemical identification of transgenic maize lines with *ZmIPT2* gene

The relative chlorophyll content, photosynthetic efficiency, and dynamic changes of cytokinin content in different growth stages of the transgenic maize lines and receptor control lines with high target gene expression were quantified.

To measure photosynthetic efficiency, photosynthetic rate was measured using a Li-6400 portable photosynthetic system (Li COR, Lincoln, NE, United States). In order to reduce the laboratory environment and the gradient in the tube of the gas exchange system, the concentration of CO₂ in the sample was set to 500 μmol^{-1} , the relative humidity in the tube was set to 65%, and the temperature of the blade in the tube was set to 28°C. Conventional methods were used to estimate outdoor *in situ* gas exchange between 8:30AM and 11:30AM. The environmental conditions in the colorimeter were as described above and were set prior to recording the gas exchange data. No less than 20 min was required to achieve a stable gas exchange measurement, i.e., photosynthetic rate change < 0.5% in 1 min.

To measure chlorophyll content, leaves [100 mg fresh weight (FW)] were incubated in 10 ml tubes containing 5 ml 80% acetone at room temperature in darkness until the tissues turned white. Total chlorophyll content was set by measuring the absorbance at 645 and 663 nm applying the formula $20.2 A_{645} + 8.02 A_{663}$ (Chory et al., 1994). The extraction and purification of isopentenyl adenine standard (IP) and zeatin in leaves to measure the dynamic changes of cytokinin content were performed according to of the protocol described by Dong et al. (2008). The cytokinin content was decided by enzyme-linked immunosorbent assay (ELISA) protocol (Wu et al., 1988). Three samples were taken from each line in each period for mixed treatment, and three technical replicates were performed for each sample.

Field experiment of transgenic maize lines

The seeds of transgenic plants and recipient inbred line C01 were planted in the transgenic experimental practice base of Northeast Agricultural University. The field experiment was performed by using the comparison method. The row spacing for each plot was 65 cm, the length was 2.0 m, and the plant spacing was 2 cm. The agronomic characteristics of each plot during maize growth were considered similar to a conventionally managed field. Three consecutive plants

TABLE 1 Amino acid compositions of *ZmIPT2*.

Types	Percentage (%)	Name	Quantity (%)
Hydrophobic amino acids	51.60	Ala	16.10
		Ile	3.10
		Leu	7.50
		Met	1.60
		Phe	3.10
		Pro	3.70
		Asn	2.20
		Val	12.70
		Trp	1.60
Hydrophilic amino acids	37.90	Asp	0.053
		Cys	0.006
		Glu	9.90
		Gly	8.70
		Ser	2.50
		Thr	5.30
		Tyr	1.20
		Gln	1.60
Basic amino acids	13.30	Arg	7.10
		His	2.80
		Lys	3.40
Acidic amino acids	15.20	Asp	5.30
		Glu	9.90

were selected in the middle row of each plot to avoid the effects of the border plants, and the identification method of agronomic traits was adopted from the Description and Data Standard of Maize Germplasm Resources. The agronomic traits in field investigation included ear length, ear diameter, axis diameter, grain length, grain width, grain thickness, bare tip length, 100-grain weight, and plot yield. Excel statistical data was used to test agronomic traits and green-holding traits, SPSS 22.0 was used for the analysis of data processing, and the *T*-test was used to test the significance of data difference.

ZmIPT2 interaction gene analysis

The amino acid sequence of the *ZmIPT2* gene was submitted to STRING⁹ to predict the proteins that interacted with the putative *ZmIPT2* protein. These gene sequences were downloaded from NCBI database. The primers for these proteins coding sequences were designed using Primer5 and are listed in Table 3. The relative expression of these interacting proteins in maize leaves at different growth stages were verified by RT-qPCR. Then, the interacting genes were further determined by bioinformatic analysis.

9 https://cn.string-db.org/cgi/input?sessionId=bAHeTVBwyfB&input_page_show_search=on

Results

The characterization and cloning of the *ZmIPT2* gene

The *ZmIPT2* gene encoded a putative protein of 322 amino acids with a predicted molecular mass of 34.49 kDa, and a predicted isoelectric point (PI) of 5.11. The amino acid sequence comparison of the *ZmIPT2* protein via ExPASy ProtParam is shown in **Table 1**. Subcellular localization predictions showed that the protein was most likely to be localized on the cytoplasm with a confidence level of 0.478 and may also be distributed in the nucleus and mitochondria but with lower confidence (**Supplementary Table 1**). The results of cross-membrane domain analysis found that the *ZmIPT2* protein did not exist in the transmembrane domain and was located in the extramembrane region with the greatest probability (probability = 1.1), indicating that the *ZmIPT2* protein did not carry out transmembrane transport, but directly exercised its function, and that it belonged to the non-transmembrane protein class (**Figure 1A**). The results of signal peptide analysis showed that there was no signal peptide in the protein sequence. The *ZmIPT2* protein was not a secretory protein and therefore could not migrate in cells (**Figure 1B**). Hydrophobic structure analysis showed that the *ZmIPT2* protein contained a typical hydrophobic region at about 16 amino acids (**Figure 1C**). Meanwhile, the *ZmIPT2* promoter indicated that it had core elements such as a TATA box and a CAAT box, as well as a large number of abiotic stress related elements referred to such as ABRE, an auxin response element, and a *cis*-acting regulatory CAT box that was involved in meristem expression (**Figure 1D**). It is speculated that the *ZmIPT2* promoter may be a stress-inducible promoter. Its predicted Secondary Structure includes alpha-helices (55.85%), random coils (31.06%), extended strands (9.90%), And β -helices (7.76%) (**Figure 1E**). Its three-dimensional homology model is a 3A8T. 1 protein sequence with a similarity of protein 40.58% (**Figure 1F**).

The CD search results indicated that the *ZmIPT2* protein contained a conserved P-loop NTPase domain (**Supplementary Figure 2**). A BLAST search of the NCBI protein database showed that the predicted amino acid sequence of *ZmIPT2* had different degrees of similarity with other known proteins. IPT protein sequences of 35 different species were downloaded, including those of *Sorghum bicolor*, *Panicum miliaceum*, *Panicum hallii*, and *Dichanthelium oligosanthes* (**Supplementary Figure 3A**). The results of the constructed evolutionary tree showed that *ZmIPT2*, *SbIPT3*, and *DoIPT3* were in the same branch, among which *ZmIPT2* and *SbIPT3* had the closest genetic relationship with an amino acid consistency of 86.18% (**Supplementary Figure 3B**). The evolutionary grouping of proteins may reveal similarities and differences in function.

RT-PCR was performed using RNA from the maize line C01 as template for the synthesis of cDNA. The amplified cDNA fragment showed a 969-bp band in a 1% agarose gel and was ligated into the pEasy-T1 Simple vector. Sequencing showed that the plasmid contained the same sequence as the candidate gene in MaizeGDB¹⁰. The reconstructed full length of the cDNA sequence, which was 1339-bp long, was identified as the sequence of the full-length *ZmIPT2* gene, including the 969 bp CDS, 155 bp 5'-UTR, and 216 bp 3'-UTR (**Supplementary Figure 4**).

Molecular detection of transgenic maize lines

Over 30 T0 independent transgenic plant lines overexpressing the *ZmIPT2* gene from *Agrobacterium tumefaciens* were established and propagated in the greenhouse. The primer pair *ZmIPT2-F/ZmIPT2-R* was shown to specifically recognize transgenic maize *ZmIPT2* by PCR amplification (**Supplementary Figures 5A,B**). Three transgenic maize lines (DNIPT2-C14, DNIPT2-C33, and DNIPT2-C34) with a single *ZmIPT2* integration event were verified by copy number analysis of the T3 generation (**Table 2**).

Quantitative real-time PCR (qRT-PCR) was used to monitor the expression of *ZmIPT2* in the leaves of transgenic lines at different growth stages (**Figures 2A,B**). The overexpressed *ZmIPT2* gene first increased and then decreased with the development of the maize plants. The relative expression level of the gene reached its peak at the R1 stage, and then decreased in the following stages. The stability of *ZmIPT2* in T2 and T3 plants was monitored by RT-PCR. No expression of the *ZmIPT2* gene was detected in receptor line C01 (**Supplementary Figures 5C,D**).

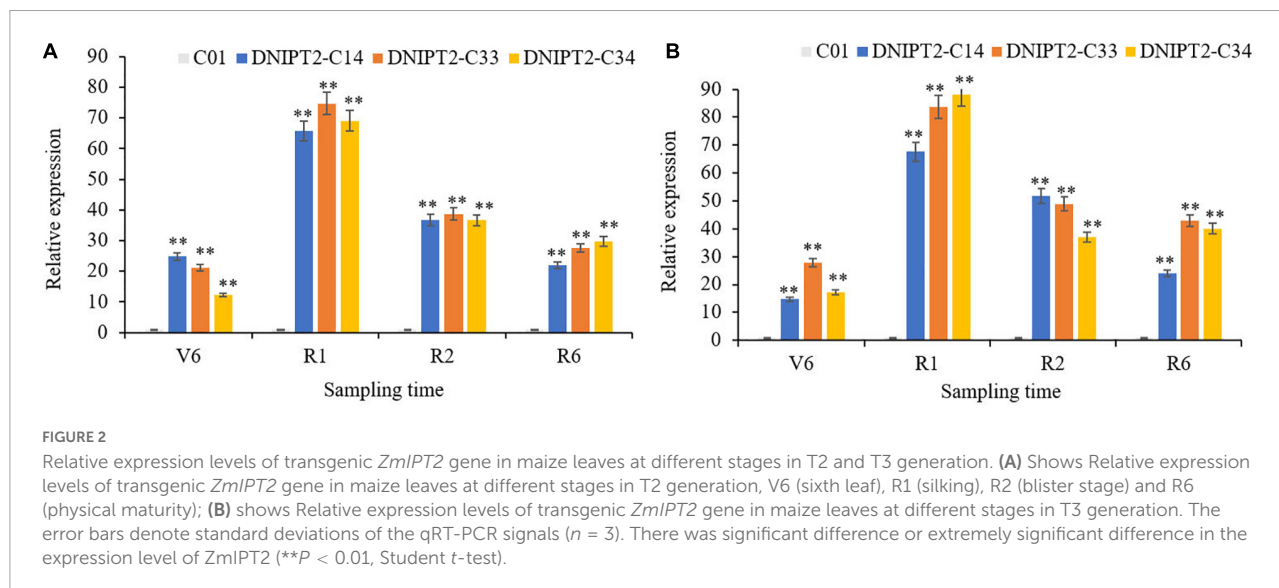
The overexpression of the *ZmIPT2* gene delayed the senescence of maize leaves

The most direct manifestations of maize leaf senescence are a decrease in photosynthetic rate and a decrease in chlorophyll content. Chlorophyll content was recorded at multiple growth stages, and the data showed that the chlorophyll levels in the transgenic lines were significantly higher in the four growth stage of the transgenic line compared to the receptor line. The transgenic plants maintained better canopy coverage and higher chlorophyll levels than recipient plants. In addition, the relative chlorophyll content of transgenic lines was the lowest at V6 (jointing stage), reached the peak at R1 (flowering

¹⁰ <http://www.maizegdb.org/>

TABLE 2 Copy number of *ZmIPT2* gene in transgenic lines.

Line	<i>Actin</i>		<i>ZmIPT2</i>		Ratio to <i>Actin</i>	Copy number
	Ct value	Amount of template	Ct value	Amount of template		
C01	25.88	0.270	28.98	0	0	0
DNIPT2-C14	23.83	0.042	29.54	0.036	0.86	1
DNIPT2-C33	20.10	0.559	24.51	0.458	0.82	1
DNIPT2-C34	20.12	0.570	25.07	0.719	1.26	1



stage), then decreased slowly for the remainder of the plants' life (Figure 3A).

It was found that among the maize plants, the transgenic lines showed a large green leaf area and obvious "stagnant green" phenomenon, while the receptor line showed a greater proportion of leaf edges withered and turned yellow (Figure 3D), thus proving that the overexpression of *ZmIPT2* gene postponed the process of leaf senescence. In addition, when compared with the receptor line, the photosynthetic rate of the three transgenic lines had significantly increased by 22.16–39.56% in the four growth stages. With the process of leaf senescence, the photosynthetic efficiency of all lines showed a downward trend after flowering. Compared with the receptor control, the photosynthetic efficiency of the transgenic lines decreased more slowly, maintained high photosynthetic efficiency for a longer time, and the plants accumulated more photosynthetic products (Figure 3B).

The CTKs content in all plants reached its highest at growth phase R1, and then began to decrease (Figure 3C). In R6, the CTKs content of the receptor line was 0.81. The CTKs content of DNIP2-C14, DNIP2-C33, and DNIP2-C34 were significantly higher than that of C01 ($P < 0.01$)

at 1.29, 1.18, and 1.07, respectively. Although the CTKs content of the transgenic line was lower than that of the first three stages, they were still significantly higher than that of the receptor line C01, indicating that the overexpression of *ZmIPT2* gene in the transgenic strain resulted in the increase of endogenous CTKs content in leaves and delayed the decline rate of CTKs.

Over-expressing of *ZmIPT2* gene increased grain yield

The important agronomic traits and yield-related traits of T2 and T3 transgenic lines were quantified (Table 3). Compared with the recipient control line, the ear length, ear thickness, grain length, grain width, and grain thickness of the transgenic lines increased significantly (Figure 4). The average plot yield of T2 and T3 transgenic plants under normal field planting conditions ranged from 12.49 to 13.05%, with the yield of the DNIP2-C14 line as the highest. Although the important agronomic traits of the three transgenic strains were improved compared with the recipient control line C01, the difference was not statistically significant, indicating that the introduction

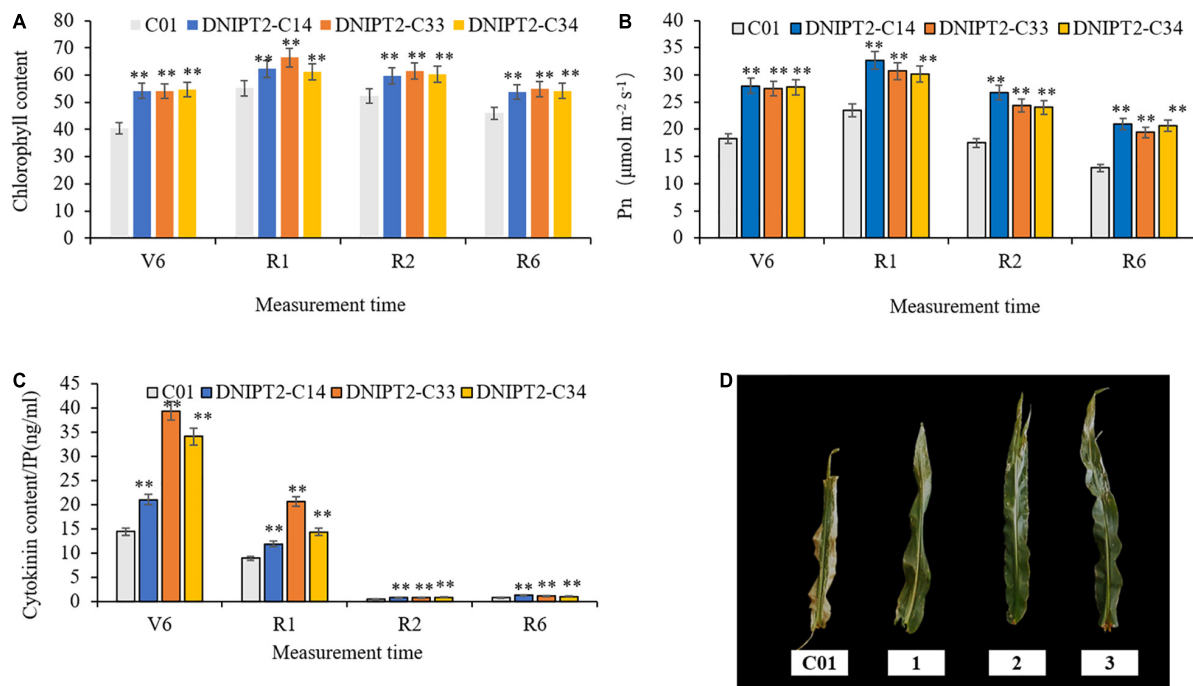


FIGURE 3

Detection of physiological and biochemical indexes of transgenic maize lines with *ZmIPT2* gene. (A) Shows the relative chlorophyll content of transgenic lines at different growth stages of maize, V6 (sixth leaf), R1 (silking), R2 (blister stage) and R6 (physical maturity); (B) shows the photosynthetic rate content of transgenic lines at different growth stages of maize, V6 (sixth leaf), R1 (silking), R2 (blister stage) and R6 (physical maturity); (C) shows the content of cytokinin (CTK) of transgenic lines at different growth stages of maize, V6 (sixth leaf), R1 (silking), R2 (blister stage) and R6 (physical maturity). (D) The comparison of leaf greenness in T3 transgenic maize lines with *ZmIPT2* gene at generation at mature stage, C01, negative control; 1, DNIP2-C14; 2, DNIP2-C33; 3, DNIP2-C34. There was significant difference or extremely significant difference in the expression level of *ZmIPT2* (** $P < 0.01$, Student *t*-test).

of *ZmIPT2* gene had no adverse effect on the growth and development of maize plants.

ZmIPT2 interaction gene analysis

Ten proteins were predicted to interact with *ZmIPT2* (Figure 5), including *GRMZM2G022904*, *GRMZM2G076936*, *GRMZM2G168681*, *GRMZM2G06198*, *GRMZM2G133082*, *GRMZM2G14772*, *GRMZM2G10828*, *GRMZM2G145029*, *GRMZM2G098569*, and *GRMZM2G027059*. *GRMZM2G022904* and *GRMZM2G076936* were cytochrome P450 superfamily proteins, which participate in the decomposition and anabolism of plant hormones. The other interacting proteins were involved in related reactions of *IPT*. qRT-PCR showed that the temporal and spatial expression patterns of *GRMZM2G168681*, *GRMZM2G147721*, and *GRMZM2G022904* in transgenic lines were consistent with those of *ZmIPT2* gene and reached their maximum at the flowering stage (Figures 5A–C). *GRMZM2G022904*, *GRMZM2G168681*, and *GRMZM2G147721* were localized in bin 7.02, 8.03, and 8.06, respectively, in regions based on B73 RefGen_v2 genome-wide data analysis. The CD-Search results indicated that their protein contains

conserved P450 superfamily and isoprenoid biosynthesis superfamily sequences (Supplementary Figure 6). In addition, *GRMZM2G022904* encoded a cytokinin hydroxylase, which can participate in the regulation of isoprenoid CK and affect the development process of plants. However, *GRMZM2G168681* and *GRMZM2G147721* have not been fully proven to interact with the *ZmIPT2* gene. The results of qRT-PCR, bioinformatics, and research reports preliminarily determined that the *GRMZM2G022904* gene may interact with *ZmIPT2* gene to participate in the molecular regulation mechanisms of anti-aging and increasing yield of maize.

Discussion

Importance of cytokinins in plant senescence

Cytokinins are known to play an important function in controlling leaf senescence and increasing grain yield of plants. There are two main approaches to regulating endogenous CTKs, which are utilized by up regulating or down regulating two kinds of genes. One of these genes encodes for *IPT*, which is

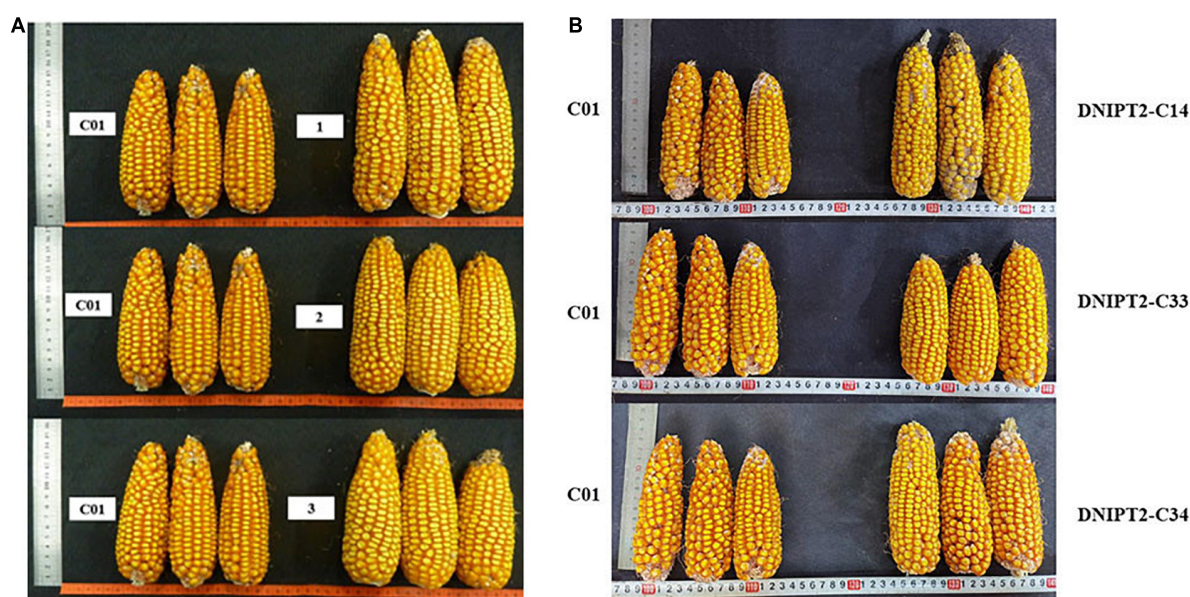


FIGURE 4

The ear and grain of transgenic maize lines with *ZmIPT2* gene and the receptor line C01. (A) The ear of T2 transgenic maize lines with *ZmIPT2* gene and the receptor line C01. (B) The ear of T3 transgenic maize lines with *ZmIPT2* gene and the receptor line C01. C01, negative control; 1, DNIPT2-C24; 2, DNIPT2-C33; 3, DNIPT2-C34.

the rate-limiting enzyme in CTK synthesis. The other is CTK oxidase/dehydrogenase, which are related to its degradation. These genes can oxidize the side chain groups of CTKs and irreversibly inactivate them.

In rice, the number of grains was increased by the accumulation of CTKs in the inflorescence meristem per panicle. Rice gene *Gn1a* encodes an oxidase/dehydrogenase (*OsCKX2*), which plays a role in the degradation of CTKs. The 11 bp deletion in the coding region of *OsCKX2* resulted in its early termination, which reduced the expression of *OsCKX2* and increased the number of grains (Ashikari et al., 2005). In wheat, the direct homolog of rice *OsCKX2*, *TaCKX6-D1*, was correlated with 1000-grain weight significantly through association analysis and linkage mapping (Zhang et al., 2012). Therefore, the regulation pathway of CTKs may play a crucial role in grain yield.

In general, leaf senescence is accompanied by the degradation of chlorophyll, proteins, and nucleic acids, as well as a decrease in photosynthetic rate. CTKs can lessen sugar accumulation, increase chlorophyll synthesis, and prolong the leaf photosynthetic period (Wu et al., 2021). Many studies have shown that the *IPT* gene plays an important role in the regulation of endogenous CTK activity required for cell division (Qin et al., 2011) and can delay leaf senescence. Guo and Gan (2011) found that the expression of the *IPT* gene at axillary bud sites was increased in *Arabidopsis* mutant MYB2, resulting in the increase of the CTKs content, and a significant delay in the aging of the whole plant (Brugiere et al., 2008).

Currently, *PSAG12:IPT*, a carrier inducing *IPT* expression by the specific promoter of leaf senescence, has been transferred to a variety of grains and cash crops such as rapeseed, wheat, tomato, cauliflower, cotton, tobacco, and corn, and has shown to significantly delay the aging of leaves and other organs, affecting plant growth and development (Guo and Gan, 2011). *PSAG12* drives *IPT* gene expression, increases cytokinin content, and delays leaf senescence. The increase of cytokinin in leaves can cause feedback into the inhibition of promoter activity and turn off the *IPT* gene, thus playing a role of self-regulation. The cytokinins synthesized by mature leaves of transgenic plants cannot be transported to other parts of the plant, thus they do not affect the normal development the plant, solving the problem of abnormal morphology of *IPT* constitutive overexpression transgenic plants. The construction of the chimeric gene *PSAG12:IPT* provides the possibility for the practical application of cytokinin biosynthesis related enzyme genes to regulate the senescence process of plant leaves (Liang et al., 2006).

Although the physiological function of CTK in delaying leaf senescence has been elucidated, the downstream molecular mechanisms of CTK in regulating leaf senescence is still not well understood. At present, it is generally believed that the CTK two-component system (TCS) pathway is involved in the regulation of CTK on leaf senescence. TCS is a multistep phosphate delivery system that includes histidine protein kinase (HKS) and downstream type A and type B response regulators (RR). The CTK response regulator (CRR) is located

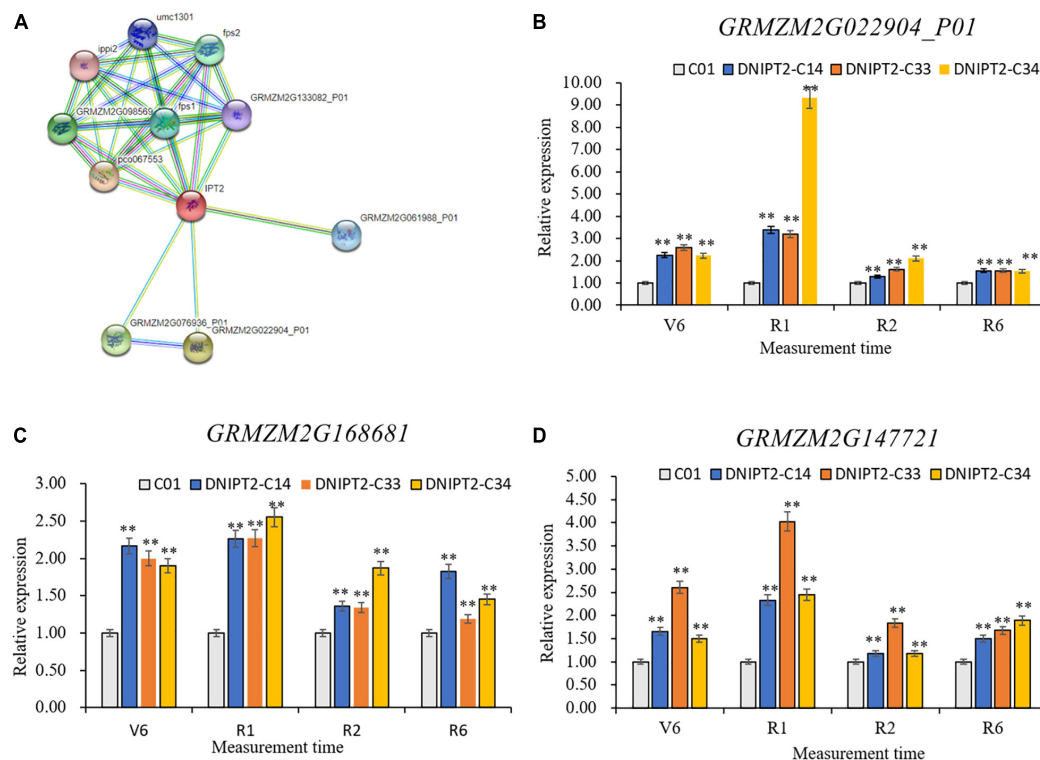


FIGURE 5

Gene analysis of *ZmIPT2* interaction. **(A)** The predictive and analysis of *ZmIPT2* interaction protein. **(B)** Relative expression levels of transgenic *GRMZM2G022904_P01* gene in maize leaves at different stages. **(C)** Relative expression levels of transgenic *GRMZM2G168681* gene in maize leaves at different stages. **(D)** Relative expression levels of transgenic *GRMZM2G147721* gene in maize leaves at different stages. There was significant difference or extremely significant difference in the expression level (** $P < 0.01$, Student *t*-test).

downstream of the signal and participates in the regulation of the physiological functions of CTK (Sakakibara, 2006). The first resolved downstream molecular mechanism of CTK regulating leaf senescence revealed that AHK3 played an important role in the CTK delay of leaf senescence. The functional deletion mutation of AHK3 delayed the CTK dependent leaf senescence and reduced the plants' sensitivity to CTK, and removed the CTK dependent phosphorylation of the downstream response factor ARR2 of type B CTK, thus regulating leaf senescence (Kim et al., 2006). AHK2/AHK3 can be used as a receptor kinase combination to participate in the accumulation and regulation of plastid coding gene transcripts. The double mutant showed phenotypes such as fewer leaf cells and low chlorophyll content, both of which significantly affected development process (Riefler et al., 2006). In addition, ARR16 could significantly induce leaf senescence through overexpression of type A CTK response factor, and the senescence process could not be inhibited by 6-BA, indicating that ARR16 was involved in the CTK regulation of leaf senescence (Ren et al., 2009).

In maize kernels, CTKs levels peaked at approximately 10 days after pollination (DAP) (Brugiére et al., 2003). Our results revealed that the chlorophyll content, CTKs, and photosynthetic rate rose substantially throughout the

development of transgenic plants. Additionally, the CTK content reached its maximum in the R1 period, which was consistent with previous reports, indicating that R1 is an important period for *ZmIPT2* to regulate CTKs, which in turn affected the later grain formation stage. Although CTKs content showed an overall decrease in the R6 period, transgenic strains still had a higher CTKs content than the receptor control. In the later stages of development, the leaves of the transgenic strains remained consistently green, and the yields of DNIPT2-C14, DNIPT2-C33, and DNIPT2-C34 were significantly higher than that of the receptor control. These results show that *ZmIPT2* gene could up-regulate the expression of CTKs content in the later stages of growth and development, improve the photosynthetic rate and chlorophyll content, and play a significant role in delaying leaf senescence, which were consistent with the results of previous studies.

Overexpression of *ZmIPT2* gene improved grain yield in maize

Genetic improvement of grain yield is of great importance to insure food security. A large number of significant agronomic

TABLE 3 The survey results of yield related traits of the transgenic maize lines with *ZmIPT2* gene.

Transgenic	Line	Ear height (cm)	Ear diameter (mm)	Cob diameter (mm)	Bald tip length (cm)	Grain length (mm)	Grain width (mm)	Grain thickness (mm)	100-grain weight (g)	Yield per plot (kg)	Yield plot increased
T2	C01	13.87 ± 0.29	46.45 ± 0.04	29.08 ± 1.62	0.00 ± 0.00	10.51 ± 1.11	9.32 ± 0.48	4.95 ± 0.58	30.41 ± 3.09	5.34 ± 0.03	
	DNIPT2-C14	15.20 ± 0.75**	45.33 ± 0.11	29.95 ± 1.20	0.57 ± 0.06*	11.49 ± 1.25*	9.11 ± 0.51	5.04 ± 0.20	30.80 ± 0.01	5.59 ± 0.01**	4.68%
	DNIPT2-C33	13.90 ± 0.70	49.17 ± 0.65*	31.70 ± 0.82	0.00 ± 0.00	11.48 ± 1.57*	8.09 ± 0.88	5.76 ± 0.62**	33.58 ± 0.00**	5.82 ± 0.00**	8.99%
	DNIPT2-C34	15.83 ± 0.64**	46.78 ± 1.74	32.23 ± 3.79*	0.47 ± 0.31	11.41 ± 0.35*	7.82 ± 0.24	5.24 ± 0.52	31.68 ± 0.00	5.66 ± 0.00**	5.99%
T3	C01	12.93 ± 0.33	42.61 ± 1.84	30.16 ± 0.66	0.00 ± 0.00	10.31 ± 0.19	8.51 ± 0.22	4.70 ± 0.46	27.32 ± 0.47	4.09 ± 0.24	
	DNIPT2-C14	17.23 ± 0.78**	44.80 ± 1.07**	27.97 ± 1.11*	0.83 ± 0.46**	10.36 ± 0.08	9.22 ± 0.35	5.29 ± 0.16	32.10 ± 0.08**	4.92 ± 0.20**	20.29%
	DNIPT2-C33	13.47 ± 0.52	45.71 ± 0.31**	31.19 ± 0.63	0.73 ± 0.23*	9.69 ± 0.33	7.92 ± 0.32	5.45 ± 0.48	31.20 ± 0.37**	4.79 ± 0.08**	17.11%
	DNIPT2-C34	14.90 ± 0.35*	48.58 ± 0.59**	33.25 ± 1.27	0.17 ± 0.17	10.20 ± 0.24	8.35 ± 0.15	5.81 ± 0.28	31.10 ± 0.78**	4.90 ± 0.08**	19.80%

Plants were grown in the molecular breeding experimental base at Northeast Agricultural University (Harbin, China) in a randomized comparative trial with three replications. The row spacing was 65 cm, the length was 2.0 m and the plant spacing was 2 cm. The results are presented as mean ± SD (*n* = 3). Asterisk (**) indicates extremely significant differences at *P* < 0.01 according to Student's *t*-test. Asterisk (*) indicates significant differences at *P* < 0.05 according to Student's *t*-test.

characteristics including yield showed continuous phenotypic variation. As a complex quantitative character, grain yield is determined many factors, such as grain number and grain weight (Melchinger et al., 1998; Yano, 2001). Various studies have proved that *IPT* gene has the function of improving crop yield and quality (Yang et al., 2017).

There were two types of *IPT*, one of which is an adenine modified tRNA, called tRNA *IPT* (EC. 2.5.1.8), which catalyzes the transfer of isopentenyl of dimethene diphosphate to the adenine residue of the precursor tRNA molecule to form a mature tRNA molecule. The modified nucleotide was located adjacent to the anti-codon, affecting the fidelity and efficiency of transcription. The other type of isopentyl transferase catalyzes the formation of IPMP, and is called *adenylate isopentenyl transferase* (*IPT: ec2.5.1.27*) which has been identified in *Agrobacterium tumefaciens* with a structure similar to tRNA *IPT*.

Studies based on ATP/ADP mutants and tRNA *IPT* showed that ATP/ADP *IPTs* are involved in the synthesis of the bulk of isopentenyladenine and *trans*-zeatin (*t-Z*) type CTK, while tRNA *IPTs* are required for *cis*-zeatin (*cis-Z*) type CTK production (Miyawaki et al., 2006). An ATP/ADP-*PpIPTs* study found that the overexpression of *PpIPT1*, *PpIPT3*, *PpIPT5*, and *PpIPT7* genes in *Arabidopsis* can increase cytokinin content in transgenic plants and improve salt resistance (Li et al., 2018). CRISPR/Cas9 targeted gene knockout *Phtheirospermum japonicum* (*Pj*)*IPT1a* inhibits parasite induced CTK response in host, revealing the important role of *PjIPT1a* in *Arabidopsis* phloem-induced CTK response (Greifenhagen et al., 2021). In the transgenic *Camellia sinensis* containing *IPT5*, the 3' UTR variant 2 (3AS2) was found to be the main transcript which participated in the regulation of tea axillary bud germination and shoot branching, providing gene resources for improving the plant type of woody plants (Zhang et al., 2021). In addition, the triple mutant of *TaIPT8-5a/5b/5d* showed a decreased *t-Z* type CTK level and reduced drought tolerance under both normal and drought conditions. In contrast, drought induced *TaIPT8* transgenic wheat plants showed stronger drought tolerance, indicating that CTK plays a beneficial role in drought tolerance by regulating the redox state of cells (Wang et al., 2022).

Previous studies have elucidated the contribution of CTKs to crop yield (Jameson and Song, 2016; Chen et al., 2020). The specific role of *IPTs* in crop yield and its role in controlling grain production should be emphasized. Many efforts have been made to modify the spatial-temporal expression of *IPTs*, using promoters with different drivers to improve crop yield. For example, the *IPT* gene is responsible for delayed foliar senescence under the control of the SAG12 promoter (Gan and Amasino, 1995), which increased plant productivity (Kant et al.,

2015). Plants that contained *IPT* driven by the SARK promoter had higher yield than the non-transgenic control (Leta et al., 2016).

Strategies for increasing yield should be focused on transgenic approaches based on the specific expression of *IPT* genes in early grain development (Sykorova et al., 2008). Cytokinin accumulation was increased in rice inflorescence meristems by decreasing the expression of the cytokinin degrading enzyme, cytokinin oxidase/dehydrogenase (*OSCKX2*), which in turn increased grain yield due to the increased number of reproductive organs (Ashikari et al., 2005). With the discovery of biosynthetic enzymes *IPT1* to *IPT10* (Irie et al., 2004; Brugiere et al., 2008; Vyroubalova et al., 2009) and degrading enzymes CKX1 to CKX12, the role of CTKs in maize kernel development has been made clearer (Morris et al., 1999; Brugiere et al., 2003; Massonneau et al., 2004; Smehilova et al., 2009; Vyroubalova et al., 2009). In *IPTs*, *IPT1*, and *IPT10* are abundantly constitutively expressed in all organs, but other *IPT* transcripts show different spatiotemporal expression patterns (Vyroubalova et al., 2009). In addition, another study showed that the modified cytokinin-degrading enzyme *Oslog15* plants had significantly increased seed setting rate, total grains, fuller grains per panicle, and 1000-grain weight under drought conditions (Wang et al., 2020). Cheikh and Jones (1994) found that in the vegetative organs of maize, the content of CTKs varied with the expression level of the related gene. When the expression level of *ZmIPT2* reached the maximum, the cytokinin also reached the maximum, which showed that the expression level of *ZmIPT2* gene in the cell core was closely related to the content of cytokinin. In maize kernels, the endosperm, especially the basal metastatic cell layer (BETL), was the main expression site of *ZmIPT2* at 8–10 days after maize pollination, and its expression in BETL and endosperm continued until late development, suggesting that this gene plays an important role in CTK biosynthesis (Brugiere et al., 2008).

In vitro studies showed that *ZmIPT2*-thymine (T) with ADP, ATP, or AMP as substrate, had higher *IPT* activity than *ZmIPT2*-cytosine (C). It was found that a favorable *ZmIPT2*-T allele was related to grain weight (Weng et al., 2013). Yang recently reported that overexpression of *IPT2* gene in maize could regulate cytokinin content. Compared to the control, the photosynthetic rate and chlorophyll content of transgenic maize were significantly increased, delaying leaf senescence and increasing yield (Yang et al., 2017). These results indicated that regulating the metabolism of CTKs in reproductive organs may be an effective way to increase crop yield by increasing the flow of assimilates from source to sink or increasing sink capacity (Weng et al., 2013).

In addition, the high yield of crops has been shown to be closely related to the amount of source material on

the plant. Leaves are the main source tissue for plants to carry out photosynthesis (Andrade et al., 2000). Therefore, crop yield can be roughly determined by the number of leaf sources, as well as the photosynthetic capacity of these leaf sources. With increasing age, leaf senescence also increases. By prolonging the functional period of leaves by delaying senescence, grain yield and plant biomass can be significantly improved (Dong et al., 2000; Thomas and Smart, 2010). Waggoner and Berger (1987) successfully used the functional period of green leaf area to express the correlation between maize leaf greenness and yield. However, some scholars believe that there is no obvious correlation between maize greenness and yield (Jiang et al., 2004; Massonneau et al., 2004; Wada and Wada, 2008).

This study discovered that the green maintenance capability of the leaves of the overexpression *ZmIPT2* strain after silking was positively correlated with the plot yield. The plot yield increased by an average of 12.49, 13.05, and 12.90%, respectively, compared to the yield of the receptor control strain. In this experiment, the *ZmIPT2* gene was transformed into maize to increase the yield, which was a rapid method when compared with other high-yield breeding methods. Three genetically stable T3 generation transgenic lines were obtained from this study. Our results showed that the overexpression of *ZmIPT2* gene could delay leaf senescence and significantly or extremely significantly improve yield traits such as ear length, ear diameter, 100 grain weight, and plot yield by significantly increasing plant CTKs content, net photosynthetic rate, and relative chlorophyll content. Among the transgenic lines, the DNIPT2-C14 line showed significantly increased yield traits when compared to the control, with a plot yield increase of 20.29% compared with the control. The field experiment was conducted for only 2 years: yield increase degree takes many years/field experiments to further confirm, so the results of this study, though promising, may be inaccurate. The three single-copy high-yield transgenic lines obtained in this will be screened out by backcross transfer method, and the target genes will be transferred into the backbone maize inbred lines for the preparation of hybrid combinations to breed new varieties of high-yield transgenic maize.

Possible regulation mechanisms of *ZmIPT2* gene

The study of *IPT*, the main regulator of plant yield, provides important insights for crop breeding such as yield improvement by way of abiotic stress tolerance. Recent studies have shown that CYP735A1 and CYP735A2 are cytochrome P450 monooxygenases (P450), which catalyze

the biosynthesis of t-Z (Takei et al., 2004). In *Arabidopsis*, it was found that the co-expression of adenosine monophenyltransferase AtIPT4 and CYP735A enabled yeast to excrete t-Z and nucleosides into the culture medium. Similarly, the *GRMZM2G022904* gene from this study is a cytochrome P450 (CYP450) superfamily protein, namely cytokinin hydroxylase, which can participate in the decomposition and anabolism of plant hormones and synthesis and metabolism of terpenoids, phenylpropanoids, alkaloids, sterols, and fatty acids in plants. Its mechanism of action in catalytic reaction is diverse and complex, and it is referred to as a universal catalyst (Schuler and Werck-Reichhart, 2003). The expression level of the *GRMZM2G022904* gene was positively correlated with the expression of *ZmIPT2* gene and positively regulated the *ZmIPT2* gene to participate in the *IPT* reaction and delay leaf senescence. However, our study only preliminarily verified that *GRMZM2G022904* was a key gene involved in the isopentenyl transferase pathway and the interaction between *ZmIPT2* and *GRMZM2G022904* by qRT-PCR. The relationship and mode of action between these two genes remains to be further studied.

Data availability statement

The original contributions presented in this study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s.

Author contributions

ZW and HD contributed to the conception and design of the work. YS, CL, and YZ performed the experiment. PG, QW, and LZ analyzed the data. YS and CL wrote the manuscript. All authors read and approved the final manuscript.

References

- An, X., Zhang, J., Liao, Y., Liu, L., Peng, D., and Wang, B. (2017). Senescence is delayed when ramie (*Boehmeria nivea* L.) is transformed with the isopentenyl transferase (*ipt*) gene under control of the SAG12 promoter. *FEBS Open Bio* 7, 636–644. doi: 10.1002/2211-5463.12191
- Andrade, F. H., Otegui, M. E., and Vega, C. (2000). Intercepted radiation at flowering and kernel number in Maize. *Agron. J.* 92, 92–97. doi: 10.2134/agronj2000.92192x
- Ashikari, M., Sakakibara, H., Lin, S., Yamamoto, T., Takashi, T., Nishimura, A., et al. (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745. doi: 10.1126/science.1113373
- Balibrea Lara, M. E., Gonzalez Garcia, M. C., Fatima, T., Ehness, R., Lee, T. K., Proels, R., et al. (2004). Extracellular invertase is an essential component of cytokinin-mediated delay of senescence. *Plant Cell* 16, 1276–1287. doi: 10.1105/tpc.018929
- Bezner, A., Faccio, P., Miralles, D. J., Abeledo, L. G., Oneto, C. D., Garibotto, M. B., et al. (2021). Stress-induced expression of IPT gene in transgenic wheat reduces grain yield penalty under drought. *J. Genet. Eng. Biotechnol.* 19, 67–84. doi: 10.1186/s43141-021-00171-w
- Brenner, M. L., and Cheikh, N. (1995). “The role of hormones in photosynthate partitioning and seed filling,” in *Plant Hormones*, ed. P. J. Davies (Dordrecht: Springer), 649–670. doi: 10.1007/978-94-009-3585-3_25
- Brugiere, N., Humbert, S., Rizzo, N., Bohn, J., and Habben, J. E. (2008). A member of the maize isopentenyl transferase gene family, *Zea mays* isopentenyl transferase 2 (*ZmIPT2*), encodes a cytokinin biosynthetic enzyme expressed during kernel development. Cytokinin biosynthesis in maize. *Plant Mol. Biol.* 67, 215–229. doi: 10.1007/s11103-008-9312-x
- Brugiere, N., Jiao, S., Hantke, S., Zinselmeier, C., Roessler, J. A., Niu, X., et al. (2003). Cytokinin oxidase gene expression in maize is localized to the vasculature,

Funding

This research was supported by the Genetically Modified Organisms Breeding Major Projects of China (2016ZX08003003).

Acknowledgments

We would like to thank the strains and carriers presented by Lixinhai, researcher of Crop Research Institute, Chinese Academy of Agricultural Sciences.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary Material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.963873/full#supplementary-material>

and is induced by cytokinins, abscisic acid, and abiotic stress. *Plant Physiol.* 132, 1228–1240. doi: 10.1104/pp.102.017707

Chang, H., Jones, M. L., Banowitz, G. M., and Clark, D. G. (2003). Overproduction of cytokinins in petunia flowers transformed with P(SAG12)-IPT delays corolla senescence and decreases sensitivity to ethylene. *Plant Physiol.* 132, 2174–2183. doi: 10.1104/pp.103.023945

Cheikh, N., and Jones, R. J. (1994). Disruption of Maize Kernel Growth and Development by Heat Stress (Role of Cytokinin/Absciscic Acid Balance). *Plant Physiol.* 106, 45–51. doi: 10.1104/pp.106.1.45

Chen, L., Zhao, J., Song, J., and Jameson, P. E. (2020). Cytokinin dehydrogenase: a genetic target for yield improvement in wheat. *Plant Biotechnol. J.* 18, 614–630. doi: 10.1111/pbi.13305

Chory, J., Reinecke, D., Sim, S. S., Washburn, T. T., and Brenner, M. (1994). A role for cytokinins in De-Etiolation in Arabidopsis (det Mutants Have an Altered Response to Cytokinins). *Plant Physiol.* 104, 339–347. doi: 10.2307/4275627

Decima Oneto, C., Otegui, M. E., Baroli, I., Beznec, A., Faccio, P., Bossio, E., et al. (2016). Water deficit stress tolerance in maize conferred by expression of an isopentenyltransferase (IPT) gene driven by a stress- and maturation-induced promoter. *J. Biotechnol.* 220, 66–77. doi: 10.1016/j.jbiotec.2016.01.014

Delatorre, C. A., Cohen, Y., Liu, L., Peleg, Z., and Blumwald, E. (2012). The regulation of the SARK promoter activity by hormones and environmental signals. *Plant Sci.* 193–194, 39–47. doi: 10.1016/j.plantsci.2012.05.005

Dong, H., Niu, Y., Li, W., and Zhang, D. (2008). Effects of cotton rootstock on endogenous cytokinins and abscisic acid in xylem sap and leaves in relation to leaf senescence. *J. Exp. Bot.* 59, 1295–1304. doi: 10.1093/jxb/ern035

Dong, S. T., Wang, K. J., and Hu, C. H. (2000). Development of canopy apparent photosynthesis among maize varieties from different Eras. *Acta Agron. Sin.* 26, 200–204. doi: 10.3321/j.issn:0496-3490.2000.02.012

Gan, S., and Amasino, R. M. (1995). Inhibition of leaf senescence by autoregulated production of cytokinin. *Science* 270, 1986–1988. doi: 10.1126/science.270.5244.1986

Ghiasi, H., Paech, C., and Dybing, C. D. (1987). Free amino acid content and metabolic activities of setting and aborting soybean ovaries. *Plant Physiol.* 85, 91–95. doi: 10.1104/pp.85.1.91

Greifenhagen, A., Braunstein, I., Pfannstiel, J., Yoshida, S., Shirasu, K., Schaller, A., et al. (2021). The Phtheirospermum japonicum isopentenyltransferase PjIPT1a regulates host cytokinin responses in Arabidopsis. *New Phytol.* 232, 1582–1590. doi: 10.1111/nph.17615

Guo, Y., and Gan, S. (2011). AtMYB2 regulates whole plant senescence by inhibiting cytokinin-mediated branching at late stages of development in Arabidopsis. *Plant Physiol.* 156, 1612–1619. doi: 10.1104/pp.111.177022

Hallmark, H. T., and Rashotte, A. M. (2020). Cytokinin isopentenyladenine and its glucoside isopentenyladenine-9G delay leaf senescence through activation of cytokinin-associated genes. *Plant Direct* 4:e00292. doi: 10.1002/pld3.292

Han, Q., Liu, R. F., Lu, L. H., and Shou, H. X. (2016). Detection of transgene copy number in transgenic soybean by real-time fluorescence quantitative PCR. *J. Nucl. Agric. Sci.* 30, 646–653. doi: 10.11869/j.issn.100-8551.2016.04.0646

Hirose, N., Takei, K., Kuroha, T., Kamada-Nobusada, T., Hayashi, H., and Sakakibara, H. (2008). Regulation of cytokinin biosynthesis, compartmentalization and translocation. *J. Exp. Bot.* 59, 75–83. doi: 10.1093/jxb/ern157

Huff, A., and Dybing, C. D. (1980). Factors affecting shedding of flowers in Soybean (*Glycine max* (L.) Merrill). *J. Exp. Bot.* 31, 751–762. doi: 10.1093/jxb/31.3.751

Irie, Y., Itokazu, N., Anjiki, N., Ishige, A., Watanabe, K., and Keung, W. M. (2004). Eugenol exhibits antidepressant-like activity in mice and induces expression of metallothionein-III in the hippocampus. *Brain Res.* 1011, 243–246. doi: 10.1016/j.brainres.2004.03.040

Jameson, P. E., and Song, J. (2016). Cytokinin: a key driver of seed yield. *J. Exp. Bot.* 67, 593–606. doi: 10.1093/jxb/erv461

Jiang, G. H., He, Y. Q., Xu, C. G., Li, X. H., and Zhang, Q. (2004). The genetic basis of stay-green in rice analyzed in a population of doubled haploid lines derived from an indica by japonica cross. *Theor. Appl. Genet.* 108, 688–698. doi: 10.1007/s00122-003-1465-z

Jin, J., Ran, Q., Zhang, X. P., Chen, M., and Chen, C. X. (2002). Investigation of the A2Σ⁺ state of CuO by laser-induced fluorescence. *Chin. Phys.* 11, 481–485. doi: 10.1088/1009-1963/11/5/313

Joshi, S., Choukimath, A., Isenegger, D., Panozzo, J., Spangenberg, G., and Kant, S. (2019). Improved wheat growth and yield by delayed leaf senescence using developmentally regulated expression of a cytokinin biosynthesis gene. *Front. Plant Sci.* 10:1285. doi: 10.3389/fpls.2019.01285

Kant, S., Burch, D., Badenhorst, P., Palanisamy, R., Mason, J., and Spangenberg, G. (2015). Regulated expression of a cytokinin biosynthesis gene IPT delays leaf senescence and improves yield under rainfed and irrigated conditions in canola (*Brassica napus* L.). *PLoS One* 10:e0116349. doi: 10.1371/journal.pone.0116349

Kieber, J. J., and Schaller, G. E. (2014). Cytokinins. *Arabidopsis Book* 12:e0168. doi: 10.1199/tab.0168

Kieber, J. J., and Schaller, G. E. (2018). Cytokinin signaling in plant development. *Development* 145:dev149344. doi: 10.1242/dev.149344

Kim, H. J., Ryu, H. J., Hong, S. H., Woo, H. R., Lim, P. O., Lee, I. C., et al. (2006). Cytokinin-mediated control of leaf longevity by AHK3 through phosphorylation of ARR2 in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 814–819. doi: 10.1073/pnas.0505150103

Kudo, T., Kiba, T., and Sakakibara, H. (2010). Metabolism and long-distance translocation of cytokinins. *J. Integr. Plant Biol.* 52, 53–60. doi: 10.1111/j.1744-7909.2010.00898.x

Kuiper, D. (1993). Sink strength: established and regulated by plant growth regulators. *Plant Cell Environ.* 16, 1025–1026. doi: 10.1111/j.1365-3040.1996.tb02052.x

Leta, T. B., Miccah, S. S., Steven, M. R., Wondyifraw, T., Charless, M., Clet, W. M., et al. (2016). Drought tolerant tropical maize (*Zea mays* L.) developed through genetic transformation with isopentenyltransferase gene. *Afr. J. Biotechnol.* 15, 2447–2464. doi: 10.5897/ajb2016.15228

Li, B., Wei, A., Song, C. X., Li, N., and Zhang, J. R. (2008). Heterologous expression of the TsVP gene improves the drought resistance of maize. *Plant Biotechnol. J.* 6, 146–159. doi: 10.1111/j.1467-7652.2007.00301.x

Li, H. L., and Zhao, D. G. (2011). IPT gene expression increased the aphid-resistant of transgenic plant in oilseed. *Mol. Plant Breed.* 9, 343–349. doi: 10.3969/mpb.009.000343

Li, M. J., Wei, Q. P., Peng, F. T., Yu, W., Luo, J. J., and Zhao, Y. F. (2018). Identification and Characterization of ATP/ADP Isopentenyltransferases (ATP/ADP PpITs) Genes in Peach. *J. Plant Growth Regul.* 38, 416–430. doi: 10.1007/s00344-018-9851-6

Li, W., Herrera-Estrella, L., and Tran, L. P. (2016). The Yin-Yang of cytokinin homeostasis and drought acclimation/adaptation. *Trends Plant Sci.* 21, 548–550. doi: 10.1016/j.tplants.2016.05.006

Liang, Q. X., Cao, G. Q., Su, M. J., and Qin, G. Y. (2006). Research progress on plant leaf senescence. *Chin. Agric. Sci. Bull.* 22, 282–285. doi: 10.3969/j.issn.1000-6850.2006.08.072

Liu, M. S., Li, H. C., Chang, Y. M., Wu, M. T., and Chen, L. F. (2011). Proteomic analysis of stress-related proteins in transgenic broccoli harboring a gene for cytokinin production during postharvest senescence. *Plant Sci.* 181, 288–299. doi: 10.1016/j.plantsci.2011.06.005

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

Massonneau, A., Houba-Herin, N., Pethe, C., Madzak, C., Falque, M., Mercy, M., et al. (2004). Maize cytokinin oxidase genes: differential expression and cloning of two new cDNAs. *J. Exp. Bot.* 55, 2549–2557. doi: 10.1093/jxb/erh274

McCabe, M. S., Garratt, L. C., Schepers, F., Jordi, W. J. R. M., Stoopen, G. M., Davelaar, E., et al. (2001). Effects of PSAG12-IPT gene expression on development and senescence in transgenic lettuce. *Plant Physiol.* 127, 505–516. doi: 10.1104/pp.010244

Melchinger, A. E., Friedrich Utz, H., and Schön, C. C. (1998). Quantitative Trait Locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL Effects. *Genetics* 149, 383–403. doi: 10.1016/1369-5266(88)80015-3

Miyawaki, K., Tarkowski, P., Matsumoto-Kitano, M., Kato, T., Sato, S., Tarkowska, D., et al. (2006). Roles of Arabidopsis ATP/ADP isopentenyltransferases and tRNA isopentenyltransferases in cytokinin biosynthesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16598–16603. doi: 10.1073/pnas.0603522103

Morris, R. O., Bilyeu, K. D., Laskey, J. G., and Cheikh, N. N. (1999). Isolation of a gene encoding a glycosylated cytokinin oxidase from maize. *Biochem. Biophys. Res. Commun.* 255, 328–333. doi: 10.1006/bbrc.1999.0199

Murray, M. G., and Thompson, W. F. (1980). Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8, 4321–4326. doi: 10.1093/nar/8.19.4321

Nagel, L. (2001). Cytokinin regulation of flower and pod Set in Soybeans (*Glycine max*(L.) Merr.). *Ann. Bot.* 88, 27–31. doi: 10.1006/anbo.2001.1423

Noh, Y. S., and Amasino, R. M. (1999). Identification of a promoter region responsible for the senescence-specific expression of SAG12. *Plant Mol. Biol.* 41, 181–194. doi: 10.1023/A:1006342412688

- Panda, B. B., Sekhar, S., Dash, S. K., Behera, L., and Shaw, B. P. (2018). Biochemical and molecular characterisation of exogenous cytokinin application on grain filling in rice. *BMC Plant Biol.* 18:89. doi: 10.1186/s12870-018-1279-4
- Peleg, Z., Reguera, M., Tumimbang, E., Walia, H., and Blumwald, E. (2011). Cytokinin-mediated source/sink modifications improve drought tolerance and increase grain yield in rice under water-stress. *Plant Biotechnol. J.* 9, 747–758. doi: 10.1111/j.1467-7652.2010.00584.x
- Peng, K. X., Zhang, W., Zhu, X. X., and Zhang, K. W. (2021). Research progress on the mechanisms of cytokinin-inhibited leaf senescence. *Plant Physiol. J.* 57, 12–18. doi: 10.13592/j.cnki.ppj.2020.0149
- Qin, H., Gu, Q., Zhang, J., Sun, L., Kuppu, S., Zhang, Y., et al. (2011). Regulated expression of an isopentenyltransferase gene (*IPT*) in peanut significantly improves drought tolerance and increases yield under field conditions. *Plant Cell Physiol.* 52, 1904–1914. doi: 10.1093/pcp/pcr125
- Reguera, M., Peleg, Z., Abdel-Tawab, Y. M., Tumimbang, E. B., Delatorre, C. A., and Blumwald, E. (2013). Stress-induced cytokinin synthesis increases drought tolerance through the coordinated regulation of carbon and nitrogen assimilation in rice. *Plant Physiol.* 163, 1609–1622. doi: 10.1104/pp.113.227702
- Ren, B., Liang, Y., Deng, Y., Chen, Q. G., Zhang, J., Yang, X. H., et al. (2009). Genome-wide comparative analysis of type-A *Arabidopsis* response regulator genes by overexpression studies reveals their diverse roles and regulatory mechanisms in cytokinin signaling. *Cell Res.* 19, 1178–1190. doi: 10.1038/cr.2009.88
- Riefler, M., Novak, O., Strnad, M., and Schmulling, T. (2006). *Arabidopsis* cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin metabolism. *Plant Cell* 18, 40–54. doi: 10.1105/tpc.105.037796
- Rivero, R. M., Gimeno, J., Van Deynze, A., Walia, H., and Blumwald, E. (2010). Enhanced cytokinin synthesis in tobacco plants expressing *PSARK::IPT* prevents the degradation of photosynthetic protein complexes during drought. *Plant Cell Physiol.* 51, 1929–1941. doi: 10.1093/pcp/pcq143
- Rivero, R. M., Kojima, M., Gepstein, A., Sakakibara, H., Mittle, R., Gepstein, S., et al. (2007). Delayed leaf senescence induces extreme drought tolerance in a flowering plant. *Proc. Natl. Acad. Sci. U.S.A.* 104, 19631–19636. doi: 10.1073/pnas.0709453104
- Robson, P. H., Donnison, I. S., Wang, K., Frame, B., Pegg, S. E., Thomas, H., et al. (2004). Leaf senescence is delayed in maize expressing the *Agrobacterium IPT* gene under the control of a novel maize senescence-enhanced promoter. *Plant Biotechnol. J.* 2, 101–102. doi: 10.1111/j.1467-7652.2004.00054.x
- Roitsch, T., and Rainer, E. (2000). Regulation of source/sink relations by cytokinins. *Plant Growth Regul.* 32, 359–367. doi: 10.1023/A:1010781500705
- Sakakibara, H. (2006). Cytokinins: activity, biosynthesis, and translocation. *Annu. Rev. Plant Biol.* 57, 431–449. doi: 10.1146/annurev.arplant.57.032905.105231
- Sambrook, J., Fritsch, F. E., and Maniatis, T. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory. *BioScience* 33, 721–722. doi: 10.2307/1309366
- Schuler, M. A., and Werck-Reichhart, D. (2003). Functional genomics of P450s. *Annu. Rev. Plant Biol.* 54, 629–667. doi: 10.1146/annurev.arplant.54.031902.134840
- Singh, R. P. (1992). Association between gene *Lr34* for leaf rust resistance and leaf tip necrosis in wheat. *Crop Sci.* 32, 874–878. doi: 10.2135/cropsci1992.0011183X003200040008x
- Smehilova, M., Galuszka, P., Bilyeu, K. D., Jaworek, P., Kowalska, M., Sebela, M., et al. (2009). Subcellular localization and biochemical comparison of cytosolic and secreted cytokinin dehydrogenase enzymes from maize. *J. Exp. Bot.* 60, 2701–2712. doi: 10.1093/jxb/erp126
- Sykorova, B., Kuresova, G., Daskalova, S., Trckova, M., Hoyerova, K., Raimanova, I., et al. (2008). Senescence-induced ectopic expression of the *A. tumefaciens ipt* gene in wheat delays leaf senescence, increases cytokinin content, nitrate influx, and nitrate reductase activity, but does not affect grain yield. *J. Exp. Bot.* 59, 3773–3787. doi: 10.1093/jxb/erm319
- Takei, K., Sakakibara, H., and Sugiyama, T. (2001). Identification of genes encoding adenylate isopentenyltransferase, a cytokinin biosynthesis enzyme, in *Arabidopsis thaliana*. *J. Biol. Chem.* 276, 26405–26410. doi: 10.1074/jbc.M102130200
- Takei, K., Yamaya, T., and Sakakibara, H. (2004). *Arabidopsis* CYP735A1 and CYP735A2 encode cytokinin hydroxylases that catalyze the biosynthesis of trans-Zeatin. *J. Biol. Chem.* 279, 41866–41872. doi: 10.1074/jbc.M406337200
- Thomas, H., and Smart, C. M. (2010). Crops that stay green. *Ann. Appl. Biol.* 123, 193–219. doi: 10.1111/j.1744-7348.1993.tb04086.x
- Vyroubalova, S., Vaclavikova, K., Tureckova, V., Novak, O., Smehilova, M., Hluska, T., et al. (2009). Characterization of new maize genes putatively involved in cytokinin metabolism and their expression during osmotic stress in relation to cytokinin levels. *Plant Physiol.* 151, 433–447. doi: 10.1104/pp.109.142489
- Wada, Y., and Wada, G. (2008). Varietal difference in leaf senescence during ripening period of advanced Indica Rice. *Jpn. J. Crop Sci.* 60, 529–536. doi: 10.1626/jcs.60.529
- Waggoner, P. E., and Berger, R. D. (1987). Defoliation, Disease, and Growth. *Phytopathology*. 77, 393–398.
- Wang, C., Wang, G., Gao, Y., Lu, G., Habben, J. E., Mao, G., et al. (2020). A cytokinin-activation enzyme-like gene improves grain yield under various field conditions in rice. *Plant Mol. Biol.* 102, 373–388. doi: 10.1007/s11103-019-00952-5
- Wang, N., Chen, J., Gao, Y., Zhou, Y. B., Chen, M., Xu, Z. S., et al. (2022). Genomic analysis of isopentenyltransferase genes and functional characterization of TaIPT8 indicates positive effects of cytokinins on drought tolerance in wheat. *Crop J.* doi: 10.1016/j.cj.2022.04.010
- Weng, J. F., Li, B., Liu, C. L., Yang, X. Y., Wang, H. W., Hao, Z. F., et al. (2013). A non-synonymous SNP within the isopentenyl transferase 2 locus is associated with kernel weight in Chinese maize inbreds (*Zea mays* L.). *BMC Plant Biol.* 13:98. doi: 10.1186/1471-2229-13-98
- Westgate, M. E., and Peterson, C. M. (1993). Cytokinin regulation of flower and pod set in Soybeans (*Glycine max*(L.) Merr.). *J. Exp. Bot.* 44, 109–117. doi: 10.1006/anbo.2001.1423
- Wiebold, W. J., and Panciera, M. T. (1990). Vascularity of Soybean racemes with altered intraceme competition. *Crop Sci.* 30, 1089–1093. doi: 10.2135/cropsci1990.0011183X003000050026x
- Wu, S. R., Chen, W. F., and Zhou, X. (1988). Enzyme linked immunosorbent assay for endogenous plant hormones. *Plant Physiol. Commun.* 5, 53–57.
- Wu, W., Du, K., Kang, X., and Wei, H. (2021). The diverse roles of cytokinins in regulating leaf development. *Hortic. Res.* 8:118. doi: 10.1038/s41438-021-00558-3
- Yang, F., Liu, C. H., Yang, X. Y., Weng, J. F., Zhou, Y., Wang, Z. H., et al. (2017). Transformation of *ZmIPT2* gene related to leaf senescence into maize and function identification. *J. Maize Sci.* 25, 45–51. doi: 10.13597/j.cnki.maize.science.20170108
- Yano, M. (2001). Genetic and molecular dissection of naturally occurring variation. *Curr. Opin. Plant Biol.* 4, 130–135. doi: 10.1016/S1369-5266(00)00148-5
- Zhang, L., Zhao, Y. L., Gao, L. F., Zhao, G. Y., Zhou, R. H., Zhang, B. S., et al. (2012). TaCKX6-D1, the ortholog of rice OsCKX2, is associated with grain weight in hexaploid wheat. *New Phytol.* 195, 574–584. doi: 10.1111/j.1469-8137.2012.04194.x
- Zhang, L. P., Li, M. H., Yan, P., Fu, J. Y., Zhang, L., Li, X., et al. (2021). A novel adenylate isopentenyltransferase 5 regulates shoot branching via the ATTTA motif in *Camellia sinensis*. *BMC Plant Biol.* 21:521. doi: 10.1186/s12870-021-03254-5



OPEN ACCESS

EDITED BY

Xia Xin,
Institute of Crop Sciences (CAAS),
China

REVIEWED BY

Jinpeng Zhang,
Institute of Crop Sciences (CAAS),
China
Guohao Han,
Institute of Genetics
and Developmental Biology (CAS),
China

*CORRESPONDENCE

Juqing Jia
jiajuqing@126.com
Cheng Liu
lch6688407@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION

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

RECEIVED 29 July 2022

ACCEPTED 12 August 2022

PUBLISHED 06 September 2022

CITATION

Zhang X, Li J, Ge Y, Guan H, Li G,
Zhang S, Wang X, Li X, Chang Z,
Zhang P, Jia J and Liu C (2022)
Molecular cytogenetic
characterization of a new
wheat-*Thinopyrum intermedium*
homoeologous group-6 chromosome
disomic substitution line with
resistance to leaf rust and stripe rust.
Front. Plant Sci. 13:1006281.
doi: 10.3389/fpls.2022.1006281

COPYRIGHT

© 2022 Zhang, Li, Ge, Guan, Li, Zhang,
Wang, Li, Chang, Zhang, Jia and Liu.
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(s)
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.

Molecular cytogenetic characterization of a new wheat-*Thinopyrum* *intermedium* homoeologous group-6 chromosome disomic substitution line with resistance to leaf rust and stripe rust

Xiaojun Zhang^{1†}, Jianbo Li^{1,2,3†}, Yudi Ge⁴, Haixia Guan³,
Guangrong Li⁵, Shuwei Zhang¹, Xiaolu Wang², Xin Li¹,
Zhijian Chang¹, Peng Zhang³, Juqing Jia^{1*} and Cheng Liu^{2*}

¹College of Agriculture, Shanxi Agricultural University, Taiyuan, Shanxi, China, ²Crop Research
Institute, Shandong Academy of Agricultural Sciences, Jinan, Shandong, China, ³Plant Breeding
Institute, School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW,
Australia, ⁴School of Life Sciences, Shanxi University, Taiyuan, Shanxi, China, ⁵School of Life
Sciences and Technology, University of Electronic Science and Technology of China, Chengdu,
Sichuan, China

Thinopyrum intermedium (JJJ⁵J⁵StSt, $2n = 6x = 42$), a member of tertiary
gene pool of hexaploid wheat (*Triticum aestivum* L., AABBDD, $2n = 6x = 42$),
provides several beneficial genes for wheat improvement. In this study,
line CH51 was developed from the BC₁F₈ progeny of a partial wheat-*Th.*
intermedium amphiploid TA18335 ($2n = 56$) and wheat cultivar (cv.) Jintai
170. Somatic metaphase chromosome counting showed that CH51 had stable
42 chromosomes. Genomic *in situ* hybridization (GISH) analysis showed that
CH51 had 40 wheat chromosomes and two *Th. intermedium* chromosomes
involving translocation between J⁵- and St-genome chromosomes. Non-
denaturing fluorescence *in situ* hybridization (ND-FISH) analysis revealed
that CH51 lacked a pair of wheat chromosome 6B. Wheat 55K SNP array
analysis verified that chromosome 6B had the highest percentage of missing
SNP loci in both CH51 and Chinese Spring (CS) nullisomic 6B-tetrasomic
6D (CS-N6BT6D) and had the highest percentage of polymorphic SNP loci
between CH51 and cv. Jintai 170. We identified that CH51 was a wheat-
Th. intermedium T6StS.6J⁵L (6B) disomic substitution line. Disease resistance
assessment showed that CH51 exhibited high levels of resistance to the
prevalent Chinese leaf rust and stripe rust races in the field. Therefore, the
newly developed line CH51 can be utilized as a potential germplasm in wheat
disease resistance breeding.

KEYWORDS

Thinopyrum intermedium, GISH, FISH, wheat 55K SNP array, disease resistance

Introduction

Hexaploid wheat (*Triticum aestivum* L., AABBDD, $2n = 6x = 42$) is one of the most essential cereal crops around the world and provides the major food source for 30% of the global population [International Wheat Genome Sequencing Consortium [IWGSC], 2014]. Wheat diseases such as rusts, powdery mildew, and *Fusarium* head blight (FHB), however, have always been major threats to wheat production in almost all the wheat growing countries. Stripe rust, caused by *Puccinia striiformis* Westend. f. sp. *tritici* (*Pst*), may cause losses up to 70% and even higher (Roelfs et al., 1992; Wellings, 2011). Leaf rust, caused by *P. tritricina* Eriks (*Pt*), is another devastating foliar disease, and can also cause severe yield reduction (McIntosh et al., 1995). Developing resistant cultivars is regarded as the most economical and effective means to control diseases. Nevertheless, because of a limited number of effective resistance genes in cultivated wheat and constantly evolving new virulent pathotypes capable of overcoming existing resistance genes in the pathogens, there is an urgent requirement to explore and utilize new resistant resources.

Thinopyrum intermedium (Host) Barkworth and D.R. Dewey (JJJ^SJ^SStSt, $2n = 6x = 42$), a perennial wild relative of hexaploid wheat, possesses many resistance genes, such as those that are resistant to stem rust, stripe rust, leaf rust, and powdery mildew pathogens (Li and Wang, 2009). Due to its high crossability with hexaploid wheat, several resistance genes have been incorporated into wheat (Li et al., 2019a). Up to now, a total of 81 leaf rust (Xu et al., 2022) and 83 stripe rust resistance genes (Li et al., 2020) have been officially named in wheat, respectively, but only *Lr38* (Friebe et al., 1993) and *Yr50* (Liu et al., 2013) were reported from *Th. intermedium*. Therefore, it is of great value to explore new *Th. intermedium* genetic resources for broadening its application in wheat biotic resistance breeding.

Substitution lines between wheat and wild relatives are regarded as the optimal bridging materials for transferring beneficial genes from wild species to cultivated wheat (Liu and Wang, 2005). Compared with addition lines, substitution lines are cytogenetically more stable (Li et al., 2019b) and preferable to produce wheat-alien translocation lines by crossing with the high pairing *ph1b* mutant (Zhang et al., 2017). Chang et al. (2010) reported that wheat-*Th. intermedium* partial amphiploid TAI8335 was highly resistant to leaf rust, stem rust, stripe rust, and powdery mildew. Later, a wheat-*Th. intermedium* disomic substitution line CH51 was selected from the BC₁F₈ progeny of TAI8335 and common wheat cultivar (cv.) Jintai 170. In this study, we used chromosome counting, genomic *in situ* hybridization (GISH), non-denaturing fluorescence *in situ* hybridization (ND-FISH), wheat 55K SNP array, and disease responses to: (1) identify the chromosome composition of CH51; (2) confirm the homoeologous relationship of *Th. intermedium* chromosomes in CH51; and (3) evaluate the responses of the

newly developed line CH51 to stripe rust, leaf rust, FHB, and powdery mildew.

Materials and methods

Plant materials

The materials used in this study included common wheat Chinese Spring (CS), Jinchun 5, Jinmai 33, Jintai 170, Mingxian 169, Nanda 2419, Taichung 29, Sumai 3, Alondra's, CS nullisomic-tetrasomic lines (CS-N6AT6D, CS-N6BT6D, and CS-N6DT6B), *Th. intermedium* (unknown origin), a partial wheat-*Th. intermedium* amphiploid TAI8335 ($2n = 8x = 56$), and its derived line CH51. TAI8335 was developed from BC₁F₈ progenies of the cross of Jinchun 5/*Th. intermedium*//Jinmai 33 (Chang et al., 2010). CH51 was selected from BC₁F₈ progenies of the cross of Jintai 170/TAI8335//Jintai 170. CS and Taichung 29 were kindly provided by Dr. Zujun Yang, University of Electronic Science and Technology of China, Chengdu, Sichuan, China. Sumai 3 and Alondra's were kindly provided by Dr. Xiue Wang, Nanjing Agricultural University, Nanjing, Jiangsu, China. All materials are maintained at Shanxi Province Key Laboratory of Crop Genetics and Gene Improvement, College of Agronomy, Shanxi Agricultural University, Taiyuan, Shanxi, China.

Genomic *in situ* hybridization analysis

Mitotic metaphase chromosomes of CH51 were analyzed by GISH according to the protocols in Zhang et al. (2001). Mitotic metaphase chromosomes were obtained from root tips and were spread according to the procedures in Lang et al. (2018). Total genomic DNA from *Pseudorogneria spicata* was used as a probe and labeled with fluorescein-12-dUTP (yellow-green fluorescence) (Enzo Life Sciences Inc., Farmingdale, NY, United States) using nick translation method. Sheared genomic DNA from CS was used as blocking DNA. Chromosomes were counterstained with propidium iodide (PI), and fluoresced red. GISH images were captured with an epifluorescence Zeiss Axioplan 2 microscope equipped with a SPOT 2.1 CCD camera (Diagnostic Instruments, Sterling Heights, MI, United States).

Non-denaturing fluorescence *in situ* hybridization analysis

Mitotic metaphase chromosomes of CH51 were further analyzed by ND-FISH according to the procedure of Fu et al. (2015). The oligonucleotide probes Oligo-pSc119.2 and Oligo-pTa535 were used to identify wheat chromosomes according to the description by Tang et al. (2014). Probe

Oligo-pSc119.2 was 5'-end labeled with 6-carboxyfluorescein (6-FAM) generating green signals, and probe Oligo-pTa535 were labeled with 6-carboxytetramethylrhodamine (TAMRA) generating red signals (Shanghai Invitrogen Biotechnology Co., Ltd., Shanghai, China). Chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI) in Vectashield mounting medium (Vector Laboratories, Burlingame, CA, United States). FISH images were captured with an Olympus BX-51 microscope equipped with a DP-70 CCD camera (Shinjuku, Tokyo, Japan).

Wheat 55K SNP array analysis

Total genomic DNA of CH51, Jintai 170, CS-N6AT6D, CS-N6BT6D, and CS-N6DT6B were extracted using the CTAB method (Chen et al., 2004), and were genotyped on the wheat 55K SNP genotyping arrays (China Golden Marker Biotechnology Company, Beijing, China). There are 53,007 microchip probes per chip, including many diploid markers. Based on CS reference genome sequence IWGSC_RefSeq_v1.0,¹ a total of 49,060 SNP marker loci had precise physical location information, evenly covering the entire wheat genome. Percentages of the same, polymorphic, or missing SNP loci in each chromosome in CH51, Jintai 170, CS-N6AT6D, CS-N6BT6D, and CS-N6DT6B were obtained by calculating the rate of the same, polymorphic, or missing SNP genotype loci number in total number of SNP loci. Microsoft Excel 2019 (Microsoft, Redmond, WA, United States) was used for data analysis and graphing.

Disease response evaluation

During the two wheat-growing seasons in 2018–2020, all materials were sown in a randomized complete block design with three replicates for evaluating their responses to stripe rust, leaf rust, powdery mildew, and *Fusarium* head blight (FHB) at the heading stage. Fifteen seeds of each line were sown in 1.5 m rows, spaced 0.25 m apart. Stripe rust was tested at Xindu Experiment Station, Sichuan Academy of Sciences, Chengdu, Sichuan, China. Leaf rust, powdery mildew, and FHB were tested at the Experimental Farm of Shanxi Agricultural University, Jinzhong, Shanxi, China.

Stripe rust responses of Jinchun 5, Jinmai 33, Jintai 170, *Th. intermedium*, TAI8335, CH51, and Taichung 29 were inoculated with a mixture of *Pst* races CYR32, CYR33, and CYR34 (1:1:1 ratio) provided by the Institute of Plant Protection, Gansu Academy of Agricultural Sciences, Lanzhou, Gansu, China. Artificial inoculations were carried out by dusting spores onto the leaves. Wheat cv. Taichung 29 was used as the susceptible control. When spores were fully developed on Taichung 29,

infection types (ITs) were recorded based on a 0–4 scale, where 0, 0; 1, 2, 3, and 4 indicated immune, highly resistant, resistant, moderately resistant, moderately susceptible, and susceptible, respectively (McIntosh et al., 1995).

Leaf rust reactions of all tested materials were recorded after being inoculated with a mixture of prevalent *Pt* races TRT, TRJ, and KHJ (1:1:1 ratio), which were collected from wheat-growing areas in northern China (Sheng et al., 2022). Inoculation method was according to Sheng et al. (2022). Wheat cv. Nanda 2419 was used as the susceptible control. ITs were recorded as 0–4 scale according to McIntosh et al. (1995).

Powdery mildew responses of all tested materials were evaluated after being inoculated with *Blumeria graminis* f. sp. *tritici* (*Bgt*) race E09 provided by the Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China. Inoculations were carried out as described by Xiang et al. (1994). When conidia were spread across the susceptible control Mingxian 169, ITs were recorded on a 0–9 scale, where 0, 0; 1–2, 3–4, 5–6, and 7–9 indicated immune, near immune, highly resistant, moderately resistant, moderately susceptible, and highly susceptible, respectively (Sheng and Duan, 1991).

Fusarium head blight (FHB) reactions of all tested materials were recorded after being inoculated with *Fusarium* pathotype F0609 provided by Dr. Xiue Wang, Nanjing Agricultural University, Nanjing, Jiangsu, China. Plants were inoculated as described by Bai et al. (1999) and Zhang et al. (2020) when a spike was just beginning to flower. Sumai 3 was used as the resistant control, and Alondra's was used as the susceptible control. Disease severity was recorded 27 days post inoculation according to a 0–4 scale, where 0, 1, 2, 3, and 4 indicated immune, resistant, moderately resistant, moderately susceptible, and susceptible, respectively (Zhang et al., 2020).

Results

Cytological characterization of CH51 using genomic *in situ* hybridization and fluorescence *in situ* hybridization analyses

Wheat-*Th. intermedium* derived line CH51 was selected from the BC₁F₈ progeny of the cross of Jintai 170/TAI8335//Jintai 170. A total of 30 CH51 seeds were germinated for chromosome counting. The result showed that the somatic metaphase chromosome number of all 30 seeds are $2n = 42$, confirming its cytogenetic stability.

Genomic *in situ* hybridization (GISH) analysis using *Ps. spicata* genomic DNA as a probe showed that CH51 had 40 wheat chromosomes and two *Th. intermedium* chromosomes displaying stronger hybridization signals along the entire short arm and at the telomeric region of the long arm

¹ <http://wheat-urgi.versailles.inra.fr/Seq-Repository/>

(Figure 1A). According to Chen et al. (1999), GISH using the diploid progenitor *Ps. spicata* as a probe could label the entire length of St-genome chromosomes, the pericentromeric and telomeric regions of J^s-genome chromosomes, and the telomeres of J-genome chromosomes, indicating that the *Thinopyrum* chromosome in CH51 involved translocation between J^s- and St-genome chromosomes. Sequential ND-FISH with probes Oligo-pSc119.2 and Oligo-pTa535 revealed that CH51 had 40 wheat chromosomes and a pair of unknown chromosomes, which substitute for the wheat chromosome 6B (Figure 1B). Therefore, we concluded that CH51 is a wheat-*Th. intermedium* T?StS.?J^sL (6B) disomic substitution line.

Wheat 55K SNP array analysis

Based on the reference genome sequence of CS, a total of 49,060 SNP loci having precise physical location information were used in the wheat SNP array analysis. Among them, a total of 46,380 and 48,288 valid SNP loci were identified in CH51 and Jintai 170, respectively (Table 1). A total of 41,186 SNP loci were common between CH51 and Jintai 170. As shown in Figure 2, chromosome 6B shared the minimum percentage of the same SNP loci (9.58%) between CH51 and Jintai 170, whereas other chromosomes shared much higher percentages of the same SNP loci ranging from 63.89% (on 6D) to 98.40% (on 4B). A total of 7,403 SNP loci were polymorphic between CH51 and Jintai 170. As shown in Figure 2, chromosome 6B had the highest percentage of polymorphic SNP loci (88.18%) between CH51 and Jintai 170, whereas other chromosomes had

lower percentages of polymorphic SNP loci ranging from 1.05% (on 7B) to 35.94% (on 6D). In addition, a total of 471 SNP loci (0.96%) were simultaneously missing in both CH51 and Jintai 170, which could not be used in the statistical analysis (Table 1). The result indicated that wheat chromosome 6B in CH51 was substituted by a pair of homoeologous group-6 chromosome from *Th. intermedium*.

To verify whether wheat chromosome 6B was absent in CH51, CS-N6BT6D, CS-N6AT6D, and CS-N6DT6B were also included in genotyping with wheat 55K SNP genotyping arrays. As shown in Supplementary Table 1, chromosome 6B in CH51 had the highest percentage of missing SNP loci of 62.82%, whereas other chromosomes had much lower percentages, ranging from 0.86% (4B) to 6.77% (6D). For CS-N6BT6D, because it lacks the wheat chromosome 6B, we speculated that it should have a highest percentage of missing SNP loci on chromosome 6B, which was confirmed by SNP array analysis that chromosome 6B in CS-N6BT6D had the highest percentage (66.98%) of missing SNP loci (Supplementary Table 1). Combined with FISH-GISH results, it was demonstrated that CH51 was a wheat-*Th. intermedium* T6StS.6J^sL (6B) disomic substitution line.

Assessment of responses to leaf rust, stripe rust, powdery mildew, and *Fusarium* head blight

At the heading stage, responses to stripe rust, leaf rust, powdery mildew, and FHB were recorded in Table 2. For stripe

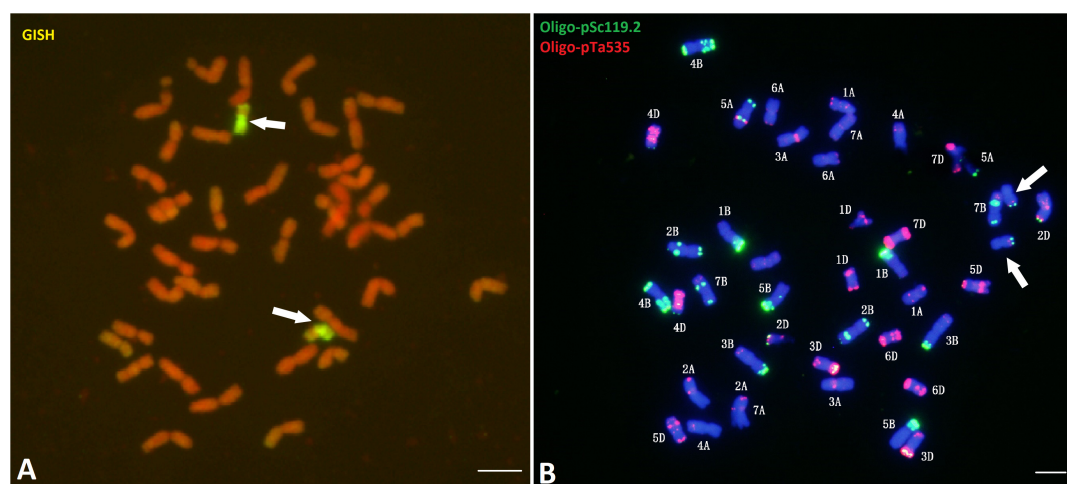
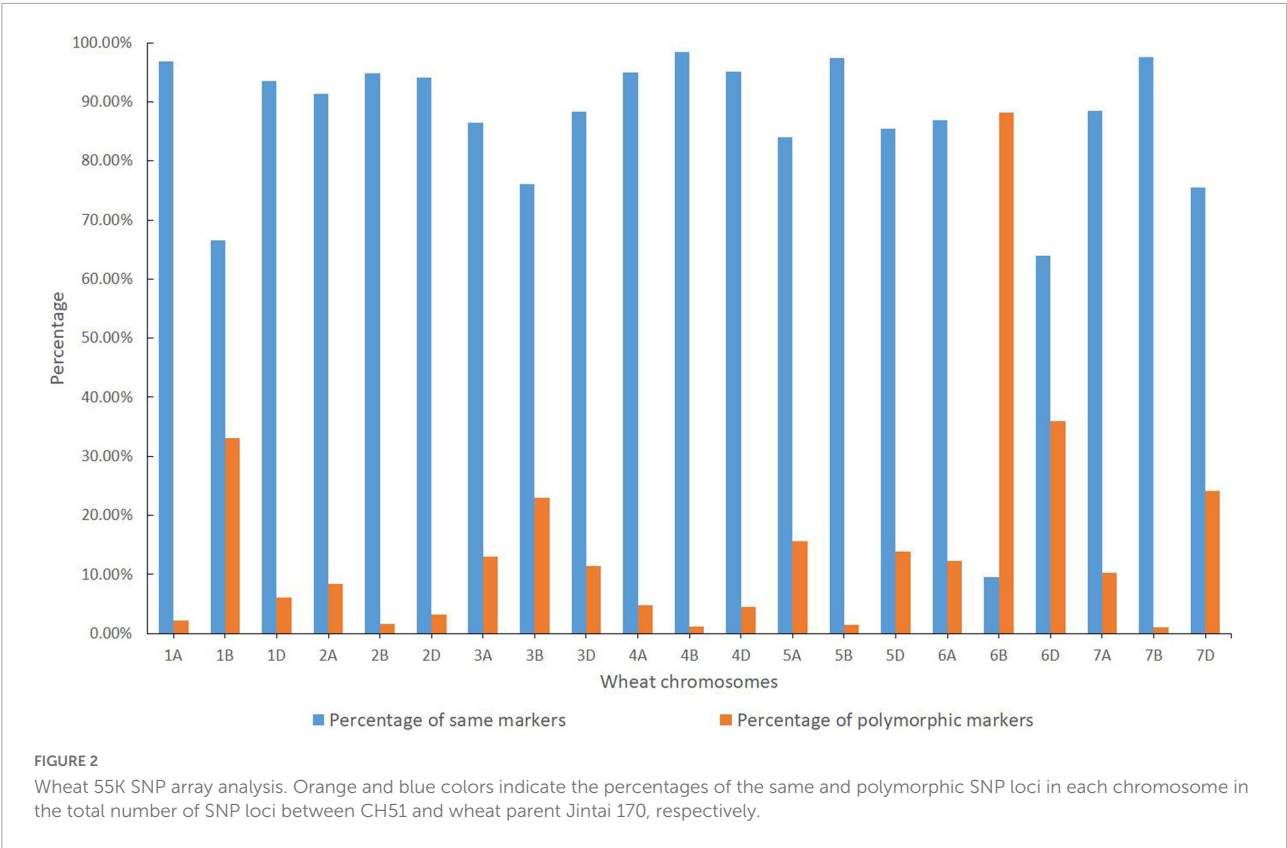


FIGURE 1

Genomic *in situ* hybridization (GISH) and Non-denaturing fluorescence *in situ* hybridization (ND-FISH) analyses of mitotic metaphase chromosomes of CH51. (A) *Pseudorogneria spicata* total genomic DNA was used as a probe (yellow-green) in GISH analysis. Chromosomes were counterstained with propidium iodide (PI) and fluoresced red. (B) Probes Oligo-pSc119.2-1 (green) and Oligo-pTa535-1 (red) were used in ND-FISH analysis. Chromosomes were counterstained with 4',6-diamidino-2-phenylindole (DAPI) and fluoresced blue. Arrows (A, B) point to the alien translocation chromosomes. Bars, 10 μ m.

TABLE 1 SNP genotyping data obtained using wheat 55K SNP arrays for CH51 and wheat parent Jintai 170.

Chromosome	No. of markers	No. of valid markers in CH51	No. of valid markers in Jintai 170	CH51 vs. Jintai 170					
				No. of same markers	Percentage of same markers	No. of polymorphic markers	Percentage of polymorphic markers	No. of simultaneous missing markers	Percentage of simultaneous missing markers
1A	2625	2581	2587	2543	96.88%	57	2.17%	25	0.95%
1B	2595	2556	2562	1728	66.59%	859	33.10%	8	0.31%
1D	2138	2107	2108	1999	93.50%	129	6.03%	10	0.47%
2A	2622	2585	2599	2394	91.30%	219	8.35%	9	0.35%
2B	2600	2486	2499	2467	94.88%	42	1.62%	91	3.50%
2D	2247	2173	2177	2115	94.13%	72	3.20%	60	2.67%
3A	2174	2130	2140	1878	86.38%	284	13.07%	12	0.55%
3B	2595	2547	2559	1974	76.07%	598	23.04%	23	0.89%
3D	1693	1600	1679	1495	88.30%	194	11.46%	4	0.24%
4A	2592	2569	2567	2460	94.91%	123	4.75%	9	0.34%
4B	2556	2534	2541	2515	98.40%	32	1.25%	9	0.35%
4D	1420	1403	1410	1350	95.07%	65	4.58%	5	0.35%
5A	2611	2580	2587	2193	83.99%	409	15.66%	9	0.35%
5B	2586	2543	2548	2519	97.41%	38	1.47%	29	1.12%
5D	1737	1716	1717	1483	85.38%	242	13.93%	12	0.69%
6A	2623	2519	2573	2279	86.89%	324	12.35%	20	0.76%
6B	2547	947	2478	244	9.58%	2246	88.18%	57	2.24%
6D	1728	1681	1690	1104	63.89%	621	35.94%	3	0.17%
7A	2579	2530	2533	2280	88.41%	266	10.31%	33	1.28%
7B	2487	2444	2441	2425	97.51%	26	1.05%	36	1.44%
7D	2305	2149	2293	1741	75.53%	557	24.16%	7	0.31%
Total	49060	46380	48288	41186	83.95%	7403	15.09%	471	0.96%



rust (Figure 3A) and leaf rust (Figure 3B), the susceptible control Taichung 29 and Nanda 2419, wheat parents Jinchun 5, Jinmai 33 and Jintai 170 were susceptible (IT 3+ or 4), whereas *Th. intermedium* and TAI8335 were immune or highly resistant (IT 0 or 0₁), and CH51 was highly resistant or resistant (IT ;1=). For powdery mildew (Supplementary Figure 1A), the susceptible control Mingxian 169, wheat parents Jinchun 5, Jinmai 33 and Jintai 170, and CH51 were highly susceptible (IT 9), whereas *Th. intermedium* and TAI8335 were immune (IT 0). For FHB (Supplementary Figure 1B), the susceptible control Alondra's, wheat parents Jinchun 5, Jinmai 33 and Jintai 170, and CH51 were highly susceptible (IT 3+ or 4), whereas TAI8335 was moderately resistant (IT 2), and the resistant control Sumai 3 was highly resistant (IT 1). Therefore, we concluded that the translocation chromosome T6StS.6J^SL in CH51 might carry genes for resistance to stripe rust and leaf rust in the field, but not resistance genes to powdery mildew and FHB.

Discussion

The homoeologous group-6 chromosomes of wild relatives of common wheat carry many desirable genes, such as higher micronutrient contents in grain and resistance to stripe rust, leaf rust, and powdery mildew. For example, Ardalani et al. (2016) reported that wheat-*Th. bessarabicum* substitution line

TABLE 2 Responses of tested materials to stripe rust (*Pst*), leaf rust (*Pt*), powdery mildew (*Bgt*), and FHB at the heading stage.

Materials	<i>Pst</i>	<i>Pt</i>	<i>Bgt</i>	FHB
	CYR32 + CYR33 + CYR34	TRT + TRJ + KHJ	E09	F0609
<i>Thinopyrum intermedium</i>	0	0	0	–
TAI8335	0 ₁	0	0	2
CH51	;1	;1=	9	3+
Jinchun 5	4	3+	9	4
Jinmai 33	4	4	9	4
Jintai 170	4	4	9	4
Mingxian 169	–	–	9	–
Nanda 2419	–	4	–	–
Taichung 29	4	–	–	–
Sumai 3	–	–	–	1
Alondra's	–	–	–	4

“–”, not tested.

DS6E^b (6D) and translocation line T6E^bS.6DL had higher iron and zinc contents than the recipient wheat cv. “Roushan” and demonstrated that the gene(s) conferring high Fe and Zn contents was located on the short arm of *Th. bessarabicum*

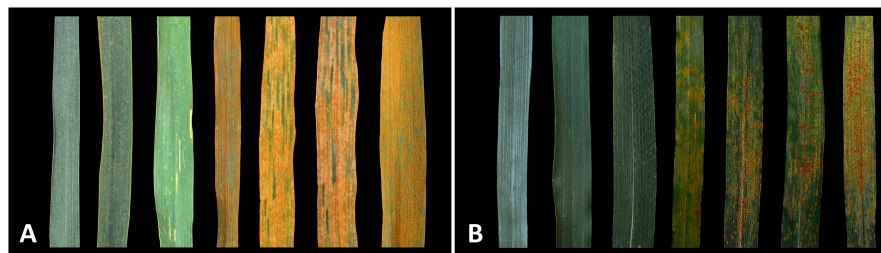


FIGURE 3

Stripe rust and leaf rust responses of tested materials at the heading stage. (A) A mixture of stripe rust races CYR32, CYR33, and CYR34 (1:1:1 ratio) were inoculated on (from left to right): *Thinopyrum intermedium*, TAI8335, CH51, Jinchun 5, Jinmai 33, Jintai 170, Taichung 29. (B) A mixture of leaf rust races TRT, TRJ, and KHJ (1:1:1 ratio) were inoculated on (from left to right): *Th. intermedium*, TAI8335, CH51, Jinchun 5, Jinmai 33, Jintai 170, Nanda 2419.

chromosome 6E^b. Song et al. (2016) revealed that the bin of fraction length (FL) 0.81–1.00 of the long arm of *Agropyron cristatum* chromosome 6P carried leaf rust resistance gene(s). The powdery mildew resistance gene *Pm21* derived from *Haynaldia villosa* is located on 6VS and encodes a CC-NBS-LRR (NLR) protein (He et al., 2018; Xing et al., 2018). Li et al. (2020) mapped a new stripe rust resistance gene *Yr83* to the bin of FL 0.73–1.00 of the long arm of *Secale cereale* chromosome 6R. Recently, Zhang et al. (2021) isolated stem rust resistance genes *Sr26* and *Sr61* from *Th. ponticum* chromosomes 6Ae#1 and 6Ae#3, respectively, which encode unrelated NLR genes and remain effective against all known *Pgt* races, including the widely virulent *Pgt* race Ug99 (TTKSK). In the present study, we identified a wheat-*Th. intermedium* T6StS.6J^SL (6B) disomic substitution line CH51, which exhibited high levels of resistance to the prevalent Chinese leaf rust and stripe rust races in the field (Figure 3).

After transferring alien chromosomes into wheat, it is important to efficiently track alien chromosome(s) in wheat-alien introgression lines. GISH is regarded as a powerful and reliable technique for determining the genomic origin, size of introgressed fragments and breakpoint positions of the introgressions (Li et al., 2020). In this study, we used GISH analysis with *Ps. spicata* genomic DNA as a probe and showed that CH51 carried a pair of *Th. intermedium* J^S-St-genome translocation chromosomes (Figure 1A). In addition, FISH is an efficient tool for the identification of wheat and alien chromosomes in wheat-alien introgression lines (Tang et al., 2014). We used FISH analysis to show that CH51 lacked a pair of wheat chromosome 6B but had a pair of *Th. intermedium* chromosomes (Figure 1B). A combination of GISH and FISH indicated that CH51 is a wheat-*Th. intermedium* T6StS.6J^SL (6B) disomic substitution line.

With the rapid development of sequencing technologies, SNP array analysis is becoming increasingly popular in high-throughput genotyping wheat and wild relatives because of its high-density loci and reasonable cost (Winfield et al., 2016). Recently, SNP arrays also play a vital role in detecting

the homoeologous relationships between wheat and alien chromosomes in wheat-alien introgression lines (Li et al., 2019b; Wang et al., 2020, 2022). In this study, results from the wheat 55K SNP array showed that chromosome 6B had the highest percentage of polymorphic SNP loci between CH51 and wheat parent Jintai 170 (Figure 2 and Table 1) and also had the highest percentage (62.82%) of missing SNP loci in CH51 (Supplementary Table 1). Combining with the cytology result, we concluded that CH51 is a wheat-*Th. intermedium* T6StS.6J^SL (6B) disomic substitution line. In addition, SNP array results also verified that the tested materials, CS-N6BT6D, CS-N6AT6D, and CS-N6DT6B, used in the current study are correct, which correspond to the highest percentage of missing SNP loci of 62.82% (6B), 67.82% (6A), 76.50% (6D), respectively (Supplementary Table 1).

In this study, TAI8335 exhibited high levels of resistance to stripe rust, leaf rust, powdery mildew, and FHB in the field. Our results showed that the translocation chromosome T6StS.6J^SL in CH51 carried resistance genes for stripe rust and leaf rust (Figure 3), but not for powdery mildew and FHB (Supplementary Figure 1). Therefore, the other six *Th. intermedium* chromosomes in TAI8335 should carry powdery mildew and FHB resistance genes and might also carry additional stripe rust and leaf rust resistance genes. For the future research, we will (1) backcross CH51 with the high pairing CS *ph1b* mutant to develop small segmental 6StS or 6J^SL translocation lines for reducing the potential linkage drag and mapping the two genes; and (2) backcross TAI8335 with common wheat for transferring powdery mildew and FHB resistance genes and/or other stripe rust and leaf rust resistance genes.

Conclusion

A wheat-*Th. intermedium* T6StS.6J^SL (6B) disomic substitution line CH51 was developed from the BC₁F₈ progeny of a partial wheat-*Th. intermedium* amphiploid TAI8335 and

common wheat cv. Jintai 170. The chromosome composition of CH51 is 14A + 12B + 14D + 2T6StS.6J^SL. CH51 exhibited high levels of resistance to the prevalent Chinese leaf rust and stripe rust races in the field. Therefore, the newly developed line CH51 can be utilized as a potential germplasm in wheat disease resistance breeding.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

Author contributions

CL, JJ, and XZ conceived and designed the research. ZC and XZ contributed to the development of the materials. ZC performed the GISH experiment. GL performed the FISH experiment. HG, JL, and JJ carried out wheat 55K SNP analysis. ZC, XZ, and YG performed powdery mildew and FHB tests. SZ, XL, and JJ performed leaf rust test. CL and XW performed stripe rust test. JL wrote the manuscript. PZ, ZC, XZ, JJ, and CL helped with analysis and edited the manuscript. All authors contributed to the manuscript and approved the submitted version.

Funding

This work was financially supported by the Shanxi Province S&T Cooperation and Exchange Special Program (202104041101017), the Youth Taishan Scholarship Project (tsqn201812123), the Key Research and Development Program of Shanxi Province (201903D221073), the Natural Science Foundation of Shanxi Province (202103021224140), and the National Key Research and Development Program of China (2016YFD0102000).

References

- Ardalani, S., Mirzaghaderi, G., and Badakhshan, H. (2016). A Robertsonian translocation from *Thinopyrum bessarabicum* into bread wheat confers high iron and zinc contents. *Plant Breed.* 135, 286–290. doi: 10.1111/pbr.12359
- Bai, G., Kolb, F. L., Shaner, G., and Domier, L. L. (1999). Amplified fragment length polymorphism markers linked to a major quantitative trait locus controlling scab resistance in wheat. *Phytopathology* 89, 343–348. doi: 10.1094/PHYTO.1999.89.4.343
- Chang, Z., Zhang, X., Yang, Z., Zhan, H., Li, X., Liu, C., et al. (2010). Characterization of a partial wheat-*Thinopyrum intermedium* amphiploid and its reaction to fungal diseases of wheat. *Hereditas* 147, 304–312.
- Chen, K., Li, F., Xu, C., Zhang, S., and Fu, C. (2004). An efficient macro-method of genomic DNA isolation from *Actinidia chinensis* leaves. *Hereditas* 26, 529–531.
- Chen, Q., Conner, R. L., Laroche, A., Fedak, G., and Thomas, J. B. (1999). Genomic origins of *Thinopyrum* chromosomes specifying resistance to wheat streak mosaic virus and its vector, *Aceria tosichella*. *Genome* 42, 289–295. doi: 10.1139/g98-131
- Friebe, B., Jiang, J., Gill, B. S., and Dyck, P. L. (1993). Radiation-induced nonhomologous wheat-*Agropyron intermedium* chromosomal translocations conferring resistance to leaf rust. *Theor. Appl. Genet.* 86, 141–149. doi: 10.1007/BF00222072

Acknowledgments

We are grateful to Dr. Ennian Yang of Sichuan Academy of Agricultural Sciences for providing the experimental field in Sichuan, Dr. Zujun Yang of University of Electronic Science and Technology of China for helping the FISH analysis, and Dr. Bernd Friebe at the Wheat Genetics Resource Center, Kansas State University for technical expertise and assistance in the GISH experiment.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1006281/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Powdery mildew and *Fusarium* head blight (FHB) responses of tested materials at the heading stage. (A) Powdery mildew race E09 was inoculated on (from left to right): *Th. intermedium*, TAI8335, CH51, Jinchun 5, Jinmai 33, Jintai 170, Mingxian 169. (B) *Fusarium* pathogen F0609 was inoculated on (from left to right): Sumai 3, Jinchun 5, Jinmai 33, TAI8335, Jintai 170, CH51, Alondra's.

- Fu, S., Chen, L., Wang, Y., Li, M., Yang, Z., Qiu, L., et al. (2015). Oligonucleotide probes for ND-FISH analysis to identify rye and wheat chromosomes. *Sci. Rep.* 5:10552. doi: 10.1038/srep10552
- He, H., Zhu, S., Zhao, R., Jiang, Z., Ji, Y., Ji, J., et al. (2018). *Pm21*, encoding a typical CC-NBS-LRR protein, confers broad-spectrum resistance to wheat powdery mildew disease. *Mol. Plant* 11, 879–882. doi: 10.1016/j.molp.2018.03.004
- International Wheat Genome Sequencing Consortium [IWGSC] (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. doi: 10.1126/science.1251788
- Lang, T., La, S., Li, B., Yu, Z., Chen, Q., Li, J., et al. (2018). Precise identification of wheat-*Thinopyrum intermedium* translocation chromosomes carrying resistance to wheat stripe rust in line Z4 and its derived progenies. *Genome* 61, 177–185. doi: 10.1139/gen-2017-0229
- Li, H., and Wang, X. (2009). *Thinopyrum ponticum* and *Th. intermedium*: The promising source of resistance of fungal and viral diseases of wheat. *J. Genet. Genomics* 36, 557–565. doi: 10.1016/S1673-8527(08)60147-2
- Li, J., Chen, Q., Zhang, P., Lang, T., Hoxha, S., Li, G., et al. (2019a). Comparative FISH and molecular identification of new stripe rust resistant wheat-*Thinopyrum intermedium* ssp. *trichophorum* introgression lines. *Crop J.* 7, 819–829. doi: 10.1016/j.cj.2019.06.001
- Li, J., Yao, X., Yang, Z., Cheng, X., Yuan, F., Liu, Y., et al. (2019b). Molecular cytogenetic characterization of a novel wheat-*Psathyrostachys huashanica* Keng 5Ns (5D) disomic substitution line with stripe rust resistance. *Mol. Breed.* 39:109. doi: 10.1007/s11032-019-1014-3
- Li, J., Dundas, I., Dong, C., Li, G., Trethowan, R., Yang, Z., et al. (2020). Identification and characterization of a new stripe rust resistance gene *Yr83* on rye chromosome 6R in wheat. *Theor. Appl. Genet.* 133, 1095–1107. doi: 10.1007/s00122-020-03534-y
- Liu, J., Chang, Z., Zhang, X., Yang, Z., Li, X., Jia, J., et al. (2013). Putative *Thinopyrum intermedium*-derived stripe rust resistance gene *Yr50* maps on wheat chromosome arm 4BL. *Theor. Appl. Genet.* 126, 265–274. doi: 10.1007/s00122-012-1979-3
- Liu, S., and Wang, H. (2005). Characterization of a wheat-*Thinopyrum intermedium* substitution line with resistance to powdery mildew. *Euphytica* 143, 229–233. doi: 10.1007/s10681-005-3862-7
- McIntosh, R. A., Wellings, C. R., and Park, R. F. (1995). *Wheat rusts: An Atlas of resistance genes*. Melbourne: CSIRO Publishing.
- Roelfs, A. P., Singh, R. P., and Saari, E. E. (1992). *Rust diseases of wheat: Concepts and methods of disease management*. Mexico: CIMMYT.
- Sheng, B. Q., and Duan, X. Y. (1991). Modification on the evaluation methods of 0-9 level of powdery mildew infection on wheat. *Biotech. J. Agric. Sci.* 9, 37–39.
- Sheng, D., Liu, M., Zhang, X., Qiao, L., Chang, L., Guo, H., et al. (2022). Characterization of leaf rust resistance in a set of wheat-*Thinopyrum* amphiploid-derived hexaploid breeding lines. *Crop Prot.* 156:105956. doi: 10.1016/j.cropro.2022.105956
- Song, L., Lu, Y., Zhang, J., Pan, C., Yang, X., Li, X., et al. (2016). Physical mapping of *Agropyron cristatum* chromosome 6P using deletion lines in common wheat background. *Theor. Appl. Genet.* 129, 1023–1034. doi: 10.1007/s00122-016-2680-8
- Tang, Z., Yang, Z., and Fu, S. (2014). Oligonucleotides replacing the roles of repetitive sequences pAs1, pSc119.2, pTa-535, pTa71, CCS1, and pAWRC.1 for FISH analysis. *J. Appl. Genet.* 55, 313–318. doi: 10.1007/s13353-014-0215-z
- Wang, S., Wang, C., Feng, X., Zhao, J., Deng, P., Wang, Y., et al. (2022). Molecular cytogenetics and development of St-chromosome-specific molecular markers of novel stripe rust resistant wheat-*Thinopyrum intermedium* and wheat-*Thinopyrum ponticum* substitution lines. *BMC Plant Biol.* 22:111. doi: 10.1186/s12870-022-03496-x
- Wang, Y., Cao, Q., Zhang, J., Wang, S., Chen, C., Wang, C., et al. (2020). Cytogenetic analysis and molecular marker development for a new wheat-*Thinopyrum ponticum* 1J⁵ (1D) disomic substitution line with resistance to stripe rust and powdery mildew. *Front. Plant Sci.* 11:1282. doi: 10.3389/fpls.2020.01282
- Wellings, C. R. (2011). Global status of stripe rust: A review of historical and current threats. *Euphytica* 179, 129–141. doi: 10.1007/s10681-011-0360-y
- Winfield, M. O., Allen, A. M., Burrridge, A. J., Barker, G. L., Benbow, H. R., Wilkinson, P. A., et al. (2016). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 14, 1195–1206. doi: 10.1111/pbi.12485
- Xiang, Q. J., Sheng, B. Q., Zhou, Y. L., Duan, X. Y., and Zhang, K. C. (1994). Analyses of resistance genes of three differential varieties to the isolates of *Blumeria graminis* f. sp. *tritici* in wheat. *Acta Agric. Boreali Sin.* 9, 94–97. doi: 10.3321/j.issn:1000-7091.1994.02.018
- Xing, L., Hu, P., Liu, J., Witek, K., Zhou, S., Xu, J., et al. (2018). *Pm21* from *Haynaldia villosa* encodes a CC-NBS-LRR protein conferring powdery mildew resistance in wheat. *Mol. Plant* 11, 874–878. doi: 10.1016/j.molp.2018.02.013
- Xu, X., Kolmer, J., Li, G., Tan, C., Carver, B. F., Bian, R., et al. (2022). Identification and characterization of the novel leaf rust resistance gene *Lr81* in wheat. *Theor. Appl. Genet.* 135, 2725–2734. doi: 10.1007/s00122-022-04145-5
- Zhang, J., Hewitt, T. C., Boshoff, W. H. P., Dundas, I., Upadhyaya, N., Li, J., et al. (2021). A recombined *Sr26* and *Sr61* disease resistance gene stack in wheat encodes unrelated NLR genes. *Nat. Commun.* 12:3378. doi: 10.1038/s41467-021-23738-0
- Zhang, P., Dundas, I. S., Xu, S. S., Friebe, B., McIntosh, R. A., and Raupp, W. J. (2017). Chromosome engineering techniques for targeted introgression of rust resistance from wild wheat relatives. *Methods Mol. Biol.* 1659, 163–172. doi: 10.1007/978-1-4939-7249-4_14
- Zhang, P., Friebe, B., Lukaszewski, A. J., and Gill, B. S. (2001). The centromere structure in Robertsonian wheat-rye translocation chromosomes indicates that centric breakage-fusion can occur at different positions within the primary constriction. *Chromosoma* 110, 335–344. doi: 10.1007/s004120100159
- Zhang, X. J., Xiao, J., Wang, H. Y., Qiao, L. Y., Li, X., Guo, H. J., et al. (2020). Evaluation of resistance to *Fusarium* head blight in *Thinopyrum*-derived wheat lines. *Acta Agron. Sin.* 46, 62–73. doi: 10.3724/SP.J.1006.2020.91015



OPEN ACCESS

EDITED BY

Chuanzhi Zhao,
Shandong Academy of Agricultural
Sciences, China

REVIEWED BY

Zulfiqar Ali,
University of Agriculture,
Pakistan
Ashish Prasad,
National Institute of Plant Genome
Research (NIPGR), India

*CORRESPONDENCE

Prashant Kumar Pandey
prashant.pandey@nrc-cnrc.gc.ca
Sateesh Kagale
sateesh.kagale@nrc-cnrc.gc.ca

SPECIALTY SECTION

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

RECEIVED 16 July 2022

ACCEPTED 19 August 2022

PUBLISHED 08 September 2022

CITATION

Pandey PK, Bhowmik P and Kagale S (2022)
Optimized methods for random and
targeted mutagenesis in field pea (*Pisum
sativum* L.).
Front. Plant Sci. 13:995542.
doi: 10.3389/fpls.2022.995542

COPYRIGHT

© 2022 Pandey, Bhowmik and Kagale. 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(s) 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.

Optimized methods for random and targeted mutagenesis in field pea (*Pisum sativum* L.)

Prashant Kumar Pandey*, Pankaj Bhowmik and
Sateesh Kagale*

Aquatic and Crop Resource Development, National Research Council Canada, Saskatoon, SK,
Canada

Field pea is an important pulse crop for its dense nutritional profile and contribution to sustainable agricultural practices. Recently, it has received extensive attention as a potential leading source of plant-based proteins. However, the adoption of peas as a mainstream source of proteins is affected by a relatively moderate protein content, anti-nutritional factors and high levels of off-flavor components that reduce protein quality. Availability of genetic variation for desirable seed quality traits is the foundation for the sustainable development of pea varieties with improved protein content and quality. Mutagenesis has been an important tool in gene functional characterization studies and creating genetic variability for crop breeding. Large-scale mutagenesis of a crop using physical and chemical agents requires diligent selection of the mutagen and optimization of its dose to increase the frequency of mutations. In this study, we present detailed optimized protocols for physical and chemical mutagenesis of pea using gamma irradiation and ethyl methanesulfonate (EMS), respectively. Gamma radiation and EMS titration kill curves were established to identify optimal doses of the two mutagenic agents. Based on germination, survival rate and growth phenotypes, a gamma radiation dose of 225 Gy and EMS concentration of 5 mM were selected as optimal dosages for mutagenesis in field pea. The presented protocol has been modified from previously established mutagenesis protocols in other crop plants. Our results indicate that the optimal mutagen dosage is genotype dependent. CRISPR/Cas-based gene editing provides a precise and rapid method for targeted genetic manipulation in plants. With the recent success of gene editing in pea using CRISPR/Cas, this innovative technology is expected to become an integral component of the gene discovery and crop improvement toolkit in pea. Here, we describe an optimized methods for targeted mutagenesis of pea protoplasts, including mesophyll protoplast extraction, PEG-mediated transformation and gene editing of a LOX gene using CRISPR/Cas system. The general strategies and methods of mutagenesis described here provide an essential resource for mutation breeding and functional genomics studies in pea. These methods also provide a foundation for similar studies in other crops.

KEYWORDS

gamma radiation, ethyl methanesulfonate, protoplasts, CRISPR/Cas (clustered regularly interspaced short palindromic repeats), legume crops, artificial mutagenesis

Introduction

Field peas, also known as dry peas, are a member of the third largest family of flowering plants Fabaceae (formerly Leguminosae). They are a rich source of prebiotic carbohydrates, protein, and micronutrients, such as Iron and Zinc. Pea seeds hold approximately 70% of the above-ground biomass protein (Ranjbar Sistani et al., 2017), which is in higher concentration and of superior quality than chickpea and cowpea (Iqbal et al., 2006). Due to its affordability, the field pea is judicially hailed as an important crop for eradicating hidden hunger in the world. The protein-rich seed profile also makes pea an important crop for the exponentially growing plant-based protein and animal-feed industry (Sim et al., 2021). Field peas are also favored in crop rotation and sustainable agricultural practices as they enrich the soil with organic matter and nitrogen, and reduce weed, disease, and pest pressures on the subsequent cereal crop (Stagnari et al., 2017). Field pea cultivation, however, faces certain challenges that must be addressed to realize its full commercialization potential.

Pea is a cool-season crop and highly sensitive to heat stress, which is predicted to increase in frequency and intensity in the future due to climate change. Similarly, the rapid spread of pea beyond its natural ecological niche has made it vulnerable to diseases and pests (Rawal and Navarro, 2019). Agronomically, there is a demand to enhance pea yields, and improve protein and starch content in the seed for commercial and nutritional purposes. There is also considerable but, thus far, unrealized potential of biofortification of mineral nutrients in the crop, which will drastically improve its nutritional profile.

Presence of genetic variation is essential for development of new varieties with improved agronomic and seed quality traits to meet the present and future challenges in pea cultivation. Breeders take advantage of natural genetic variation that is generated by spontaneous mutations. Several natural pea germplasm collections have been established, including the USDA and European pea germplasm collections (Jing et al., 2012; Al Bari et al., 2021). The analysis of genetic polymorphisms in these diverse germplasm pools is currently underway (Burstin et al., 2015; Pandey et al., 2021; Shirasawa et al., 2021; Crosta et al., 2022) to facilitate the association of key phenotypes with the identified genetic polymorphisms. However, continuous selection of only yield-related traits among crosses of genetically related pea cultivars has led to narrowing of the genetic base of the crop, particularly affecting the historically neglected protein-related traits. Therefore, complementing traditional breeding approaches with well-functioning supplementary tools, such as mutation breeding and genome editing, is required to create and exploit genetic variation in agronomic and seed quality traits, thereby removing the barriers to the adoption of pea as a leading source of plant-based proteins.

The introduction of novel genetic variation through artificial mutagenesis has accelerated the development of new varieties, functional characterization of genes and the establishment of gene-trait relationships in several crop species (Sikora et al., 2012;

Zhu et al., 2012). Random mutagenesis can be carried out using chemicals [such as N-methyl-N-nitrosourea (MNU), sodium azide, ethyl methanesulfonate (EMS)] or physical mutagen agents (gamma rays, X-rays, UV-rays, and fast-moving neutrons), and have been widely used for introduction of genetic variation in model and crop species. These techniques are especially important for the introduction of genetic variation in crops recalcitrant to Agrobacterium-mediated genetic modifications, such as peas (Dalmais et al., 2008). Over the last six decades, 3,402 mutagenized varieties from 257 economically important species have been registered with United Nations International Atomic Energy Agency and FAO.¹

Mutagenesis can cause a range of genetic variations, including single nucleotide polymorphisms SNPs; leading to either transition ($A \leftrightarrow T/G \leftrightarrow C$), transversion ($A \leftrightarrow G/T \leftrightarrow C$), nucleotide insertion or deletion events (indels), chromosomal breaks, and/or chromosomal re-arrangements. Chromosomal breaks or rearrangements can affect multiple genes simultaneously, and are likely to induce the most drastic phenotypes. Smaller effects, such as SNPs and short indels could produce synonymous (non-consequential) or non-synonymous (consequential) effects, with the latter possibly leading to alteration of regulatory regions, premature termination of a gene (nonsense mutation), or alteration of the reading frame (frameshift mutation). The high-energy radiations in physical mutagenesis can cause single-strand nicks or double-strand breaks in the genetic material, predominantly leading to large chromosomal breaks and chromosomal rearrangements, along with SNPs. Chemical mutagenesis, on the other hand, generally causes chemical alterations in specific nucleotides that cellular proof-reading machinery misreads and modifies causing SNPs. For example, EMS causes base changes of $G \rightarrow A$ and $C \rightarrow T$.

Induced mutagenesis has found major application in crop improvement programs (mutation breeding), where mutated elite/commercial crop varieties could directly be screened for traits of interest. Unlike genetic modification using transgenic approaches, the mutation breeding approach is accepted by major crop regulatory authorities worldwide. Artificial mutagenesis has also revolutionized basic research by facilitating functional characterization of genes and alleles. Large-scale genotyping and sequencing of such mutagenized populations has led to the establishment of public repositories of genetic variants for several plant species (Dalmais et al., 2008; Chen et al., 2012; Schreiber et al., 2019) which can be used for gene discovery and functional genomics studies (Clemente et al., 2015). Mutagenesis approaches have also been used in generating the much-needed diversity in vegetatively or asexually propagated crops, such as bananas, roots and tubers, ornamental plants (Hernández-Muñoz et al., 2019; Wang et al., 2021), and crops recalcitrant to Agrobacterium-mediated genetic modifications (Dalmais et al., 2008).

¹ <https://mvd.iaea.org/>

While artificial mutagenesis has been tremendously useful in producing improved cultivars and new knowledge of gene functions (Sikora et al., 2012; Nasti and Voytas, 2021), there are technical and practical constraints associated with its application. For example, random mutagenesis is likely to show limited success with polygenic traits, especially in complex polyploid genomes. In such cases, only those alleles which have the highest contribution to a trait are likely to be identified, while the effect of other contributing alleles could be diluted. Also in the case of polyploid crops, the non-mutated homoeologs might compensate the function of the mutated gene/allele, thus masking the phenotypic effect. The random mutagenesis techniques might also show limited success toward epigenetically-regulated phenotypes. This is especially true for chemical mutagenesis where SNPs are less likely to drastically affect the epigenome landscape. However, deletion of an epigenetically modified region using physical mutagenesis is more likely to be useful for such traits. Homozygous mutations often provide clearer read-outs as compared to heterozygous mutations. The fixing of a mutation in subsequent generations of self-pollinating crops, such as pea, makes them easier to work with as compared to the cross-pollinating species. Another practical constraint with the technique is the cost associated with the identification of causal mutations through sequencing techniques.

Artificial mutagenesis has been successfully performed on several plant parts. Mature seeds are the most convenient and commonly used material for mutagenesis as they can be generated in large amounts, and are easy to handle and store. However, other plant materials, such as plant calli, immature inflorescence, isolated immature embryos, anthers, pollen grains, and vegetative propagules have also been used (Joseph et al., 2004; Yang et al., 2004; Ookawa et al., 2014; Serrat et al., 2014; Perera et al., 2015).

The choice between physical or chemical mutagen depends on the aim of the study, the plant material to be mutagenized, and the availability of infrastructure. If a study aims at identifying genes or key nucleotides associated with a trait of interest, then chemical mutagenesis might be preferable. However, if the study aims to identify a genomic locus associated with the trait of interest, especially in an uncharacterized genomic region, then physical mutagenesis might be preferable. Generally, plant seeds are amenable to both physical and chemical mutagenesis, however, large seeds might benefit from high and uniform penetrability of physical mutagens, such as gamma rays or fast-moving neutrons (Oladosu et al., 2016). The high penetrability characteristic of physical mutagens is also associated with the high reproducibility of mutation density (Zheng et al., 2020). However, physical mutagenesis requires access to specialized infrastructure and highly trained personnel to handle the equipment, while chemical mutagenesis could be conducted in a basic laboratory set-up with some precautions.

The optimal dosage of mutagen needs to be quantified diligently. The dosage depends on the plant material and the aim of the study. For a polygenic trait, where multiple genes contribute to a phenotype, high density of mutations might be desirable to allow mutations in multiple genes affecting the phenotype. On the

other hand, high density mutations might hinder isolation of specific mutations causing the phenotype due to the background noise. The mutation density is only partially correlated with the mutagen dosage, as high dosages also increase the likelihood of lethality. Therefore, a pilot study to optimize the dose of mutagens is highly recommended.

Here we provide a step-by-step modified protocol for gamma irradiation and EMS treatment of field pea seeds to determine optimal dosages for mutagenesis. We tested several conditions (dosages and time of exposure) on a small batch of field pea seeds (~200 seeds per condition), followed by an analysis of biological effectiveness of the mutagen by measuring germination rate, survival rate and various other growth phenotypes. This analysis helped in the identification of gamma radiation and EMS dosages that allow 50% germination and survival rate (generally termed as lethal dose 50, LD₅₀) of field pea seedlings.

Compared to the imprecise and less efficient random mutagenesis, CRISPR/Cas-based genome editing approaches provide a faster way of targeting individual genes or gene families in crop plants. While CRISPR/Cas system appears to work universally, the efficiency of targeted mutagenesis in pea varies greatly because of its recalcitrant nature, and low frequency of transformation and plant regeneration. Continuous efforts are being made by the research community to improve transformation and gene editing in pea (Bhowmik et al., 2021). The first report of successful generation of gene edited pea was recently published (Li et al., 2022). However, to best of our knowledge, CRISPR/Cas-mediated targeted mutagenesis has not been reported in pea protoplasts. Here, we provide the detailed description of methods and procedures for creating targeted edits in pea protoplasts using the CRISPR/Cas system.

Materials

Mutagenesis in pea

Plant material and growth conditions

Seeds of the field pea genotype CDC Amarillo (Warkentin et al., 2014) obtained from Galloway Seeds (Alberta, Canada) were used for mutagenesis experiments described in this study. After mutagenesis, dry field pea seeds were carefully sown at 3–5 cm depth using blunt-tipped forceps in water-saturated soil mixture (Sunshine mix # 8, SunGro Horticulture, Canada) in 48 pot seeding trays. Each tray was labeled with the mutagen and dose to which the seeds were subjected. The trays were maintained in a growth chamber at 24°C, 16 h/8 h (day/ night), 60% humidity.

Chemicals and reagents

1. Ethyl methanesulfonate (Sigma-Aldrich Catalogue number: M0880). EMS (MW 124.16 g/mol) is supplied in liquid form at density of 1.206 g/ml (at 20°C). The working

solutions are prepared in water, as described in Table 1. It is important to thoroughly mix the EMS stock in the working solution.

2. Ultra-pure water (to prepare EMS solution and washing of seeds).
3. Sodium thiosulfate (10% w/v in water).

Personal protective equipment

1. Disposable chemical resistant long-sleeved lab coats or coveralls.
2. 3 M half-face reusable respirators (3M 7000) with particulate filter P100 (3M 7093).
3. Chemical resistant/industrial nitrile gloves with an extended cuff.
4. Lab safety goggles.
5. Full face shield.

Consumables

1. Mesh bags for seeds.
2. Glass beakers (500 ml for EMS treatment, 2 liters for seed washes).
3. Glass stirrer.
4. Long-handled tongs.
5. Magnetic beads (optional).
6. Magnetic stirrers (optional).
7. Pipettes.
8. Filtered pipette tips.
9. Measuring cylinders.
10. Large weighing boats.

Gamma irradiation of pea seeds

A Gammacell 220 (SN:236) with a ^{60}Co source was used for γ -irradiation. At the time of irradiation, the central absorbed dose rate was calculated to be 1.19 Gy/min, and the calibration was confirmed by alanine dosimetry in 2021.

TABLE 1 Preparation of EMS solutions (200 ml) for mutagenesis.

Concentration of working solution (200 ml) (mm)	EMS stock (μl)
5	81.5
10	163
15	303
20	401
25	508
30	606

The dose rate was calculated based on the original cell calibration (5,180 Gy/h) using the half-life of ^{60}Co with first order decay:

$$N(t) = N_0 \left(\frac{1}{2} \right)^{\frac{t}{t_{1/2}}}$$

where $N(t)$ quantity of ^{60}Co remaining, N_0 is the initial quantity of ^{60}Co , t is the time elapsed and $t_{1/2}$ is the half-life of ^{60}Co , 5.271 yrs.

Targeted mutagenesis in pea protoplasts using CRISPR/Cas system

Plant protoplasts provide a unique single cell system for conducting cell-based experiments using different molecular, cellular or genomic tools (Davey et al., 2005). With single cell analysis, it is also easier to detect a low frequency of mutated cells than it is with a pooled population. Although many successful reports of CRISPR/Cas mediated gene editing have now been published, effective delivery of genome editing components and early detection of mutation, two keys aspects to achieving high efficiency gene editing, still remain challenging for most of the pulse crops. Successful protoplast isolation depends on careful selection and optimization of several factors including the age of the leaf tissues, duration of enzyme incubation, enzyme concentration, gentle agitation and nature of the osmoticum (Sinha et al., 2003).

Plant material

Donor plant leaves of the CDC Meadow pea variety (a cultivar from the Crop Development Centre, University of Saskatchewan; Warkentin et al., 2007) were used for protoplasts isolation.

Chemicals and reagents

Cell wall enzyme solution (50 ml)

0.6 M Mannitol 5.46 g.
MES 10 mm 1 ml of a 0.5 M Stock.
Incubate at 70° for 5 min.

1.5% cellulase R-10 0.75 g

0.75% Macerozyme R-10 0.37 g.
Incubate at 55° for 10 min.
CaCl₂ (10 mm) 0.5 ml of a 1 M Stock.
BSA (0.1%) 1 ml of a 5% Stock.
Bring up to 50 ml, filter sterilize with syringe filter.

W5 solution (500 ml)

MES (2 mm) 2 ml of a 0.5 M Stock.
NaCl (154 mm) 77 ml of a 1 M Stock.
CaCl₂ (125 mm) 62.5 ml of a 1 M Stock.
KCl (5 mm) 2.5 ml of a 1 M Stock.
Bring up to 500 ml, filter sterilize.

MMG solution (50 ml)

MES (4 mm) 0.4 ml of a 0.5 M Stock.
0.4 M Mannitol 3.64 g.
MgCl₂ (125 mm) 0.75 ml of a 1 M Stock.
Bring up to 50 ml, filter sterilize.

Peg solution (10 ml)

40% PEG 4000 4.0 g.
0.2 M Mannitol 0.36 g.
CaCl₂ (1 M) 1 ml of a 1 M Stock.
Bring up to 10 ml.

0.55M sucrose (250 ml)

Sucrose 47.07 g.
Bring up to 250 ml. Filter sterilize.

0.6M manitol (500 ml)

Mannitol 54.65 g.
Bring up to 500 ml. Filter sterilize.

0.5M MES-KOH (50 ml)

MES 4.88 g.
pH to 5.7 with KOH (uses quite a bit). Bring up to 50 ml.

Consumables and equipments

1. Surgical blades.
2. 0.22 µm syringe sterilization filter.
3. 15 ml round bottom centrifuge tube.
4. 40 µm nylon mesh.
5. Autoclave.
6. 35 mm Cell culture dish.
7. Pipettes.
8. Filtered pipette tips.
9. Measuring cylinders.
10. Weighing boats.
11. Lamina flow hood.
12. Microcentrifuge and swing-out rotor centrifuge.
13. Plant DNA purification kit.
14. Plasmid purification kit.
15. Thermal Cycler.
16. NanoDrop™ Spectrophotometer.

Methods

Pre-treatment conditions for chemical and physical mutagenesis

1. Mutagenesis should be performed on a genetically homogeneous seed material to minimize polymorphism in the starting material.

2. Healthy looking seeds of relatively similar size should be selected to ensure uniformity of mutagenesis.
3. Fume hood should be used for chemical mutagenesis. The place of work should be decluttered to minimize spills and droplet contaminations.

Chemical mutagenesis

EMS, the most widely used chemical mutagen in plants, was selected for chemical mutagenesis of field pea. To determine the optimal dose of EMS, kill curve was established at different concentrations. In total six different EMS concentrations (5, 10, 15, 20, 25, and 30 mm) with a constant incubation period of 18 h were used. The selection of the range of EMS concentrations was guided by previous work on closely related species (Rawal and Navarro, 2019).

Ems treatment procedure

1. Aliquots of 200 healthy-looking seeds each for control and the selected EMS doses were prepared in disposable, pre-labeled nylon mesh bags.
2. For each treatment, pea seed bags were fully immersed in 200 ml of EMS solution for 18 h at room temperature. The volume of EMS solution should be sufficient to allow complete submersion of the seeds. The seeds could optionally be stirred in the solution using a magnetic stirrer or beads. EMS solution has a relatively short half-life of 48.5 h at 25°C. While the EMS solutions could be prepared 12–16 h before treatment, we recommend that they are freshly prepared to avoid inconsistency of the mutagen potency in replicates.
3. After 18 h of incubation, an equal volume of 10% sodium thiosulphate (w/v) solution was added to the EMS solution, gently stirred, and allowed to stand for 10 min. This aids in neutralizing the EMS solution and stopping the mutagenesis process. This step is important to ensure the reproducibility of results, especially when multiple seed batches are handled simultaneously.
4. Using long-handled tongs, the seeds bags were lifted from the neutralized EMS solution allowing excess liquid to drain off, and each seed bag was washed separately in 1000 ml water (in a 2000 ml beaker to avoid overflow) for a minimum of 10 min while occasionally stirring the solution gently to avoid splashes. The washing step was repeated five times.
5. Next, the EMS treated seeds were transferred to a wide pre-labeled weighing boat with blotting sheets and allowed to dry for at least 24 h.
6. Dried seeds were immediately planted into trays in the greenhouse or a growth chamber. Optionally the seeds could be stored at 4°C for a short duration before planting.

Precautionary measures in handling EMS

EMS is a powerful mutagen and extremely hazardous, therefore, special care (as described below) must be taken while handling the reagent or any contaminated material.

1. The personnel handling EMS must wear double layers of gloves with a long breakthrough time to prevent contact with the mutagen. Personnel should also wear appropriate personal protective equipment (PPE) listed in the materials section.
2. All materials contaminated with EMS should be collected and categorized separately as washable or disposable. Appropriate collection bags or buckets should be prepared before starting the experiment and EMS contaminated materials should be placed in designated buckets immediately after use. We recommend the use of three separate containers: for EMS contaminated glassware (washable), contaminated plastic ware and PPE (disposable solid waste), EMS liquid waste.
3. The contaminated glassware should be soaked in 10% sodium thiosulphate (w/v) solution for 12–15 h, followed by washing with water (using wash bottles) to neutralize and remove EMS. This soaking and washing solution should be discarded in the liquid EMS waste container. The glassware can then be cleaned with other labware.
4. The liquid EMS waste should be neutralized using 10% sodium thiosulphate (w/v) solution (1:1 dilution) and transferred to EMS waste container.
5. The efficiency of EMS mutagenesis is influenced by the temperature of the room and prepared solutions, and also the presence of catalytic ions (Zn^{2+} , Cu^{2+}). Therefore, the room/fume hood and the solutions to be used should be maintained at 25°C and deionized water should be used to ensure reproducibility of the procedure.
6. Magnetic stirrers or beads could be used to stir the EMS solution during overnight incubation of seeds. However, care must be taken to maintain low stirring speeds to avoid droplet dispersion.
7. It is advised to close the fume hood sash during the overnight incubation or when the fume hood is not actively monitored by the personnel. If the fume hood is in a shared lab, cautionary signage is highly recommended to avoid accidental contaminations.
8. It is essential to completely dry the seeds after EMS treatment, failure to do so could lead to degradation of seeds. Therefore, the use of large weighing boats to allow uniform dispersion and aeration of the seeds is recommended.
9. While handling EMS-mutagenized seeds, personnel should wear appropriate PPE and gloves. The growth chamber used for EMS-mutagenized seeds should be clearly labeled with cautionary signs.

Physical mutagenesis

For field pea seeds, we chose 150, 200, 250, 300, 350, and 400 Gy doses with Cobalt-60 as the source for gamma radiations.

1. Approximately 200 seeds were aliquoted in nylon mesh bags for each dose and one for the control sample.
2. Seed bags were incubated in a vacuum desiccator over either silica, zeolite desiccant beads, or glycerol (60% v/v in water) for 1 week to equilibrate moisture content in all seeds.
3. Seed bags were then exposed to the desired gamma radiation dose using Gammacell 220 (SN:236) at the Saskatchewan Structural Sciences Centre, University of Saskatchewan, Saskatoon, Canada. This step was performed by a trained operator at the facility.
4. The radiation dose depends on the time of exposure to the radiation source and the dose rate. The exposure time is calculated as:

$$\text{Exposure time} = \frac{\text{Dose}}{\text{Dose rate}}$$

Gamma-irradiated seeds were immediately planted into trays in the greenhouse or a growth chamber or stored at 4°C for a short duration before planting.

Precautionary measures for gamma irradiation

1. The efficiency and reproducibility of physical mutagenesis are highly dependent on the moisture content of the starting material. Therefore, the steps involving equilibration of moisture content in seeds should not be skipped.
2. The efficiency and reproducibility of physical mutagenesis also depends on foreign material, such as dust or fiber. Therefore, care should be taken to minimize contamination.
3. Mutagenesis should not be performed on seeds infected with pathogens.

Assessment of seed viability after EMS treatment and gamma irradiation

The assessment of seed viability to determine optimal doses of EMS and gamma radiation for pea mutagenesis was carried out based on the germination rate of treated seeds, which could be assessed by allowing the seeds to germinate on water-soaked Whatman sheets in a Petri dish. Although this is the most convenient method, it is also prone to microbial contaminations under prolonged incubation periods often caused by delayed germination of mutagenized seeds. Hence in our analysis, we analyzed the optimal dose on soil-grown plants as it allowed us to phenotype several post-germination traits, such as growth rate (time taken for the appearance of two- and four-leaves), and chlorosis. The control (non-mutagenized seeds) were also seeded at the same time for comparative analysis of phenotypes. To assess the effect of different EMS concentration and gamma radiation doses on seed viability and survival of pea seedlings, the germination of seeds, greening of cotyledon, and the appearance

of second and fourth leaf was recorded for each of the treatment conditions.

Gene editing in pea protoplasts

Mesophyll protoplasts isolation and purification

1. Pea seeds were surface sterilized with 75% ethanol for 1 min in a 15 ml centrifuge tube. After removing the ethanol using pipette, 10 ml of bleach solution was added and the tube was placed on a shaker for 15 min.
2. After 15 min, seeds were washed with sterilized water for five times and placed on germination media in a tissue culture cabinet.
3. After 3 weeks of subculturing leaf tissues were surface sterilized with 70% ethanol and then, using a new scalpel blade, cut into 0.5–1 mm pieces.
4. The pieces were quickly transferred to a plate containing 10 ml of 0.6 M mannitol and placed on a shaker for 10 min at 50 rpm in the dark.
5. The mannitol solution was removed and 10 ml of warm enzyme solution (0.6 M Mannitol, 10 mM MES, 1.5% Cellulase R-10, 0.75% Macerozyme R-10, 10 mM CaCl_2 , 0.1% BSA) was added. The plate went back onto the shaker at 50 rpm for 4 h in the dark.
6. Protoplasts were then transferred through a 40 μm filter, washed with W5 and layered overtop of 5 ml of 0.55 M sucrose in a round bottom culture tube.
7. Protoplasts were centrifuged for 10 min at 100 g in a swinging bucket without using brakes on at 4°C. The green intermediate layer was transferred to a sterile round bottom culture tube then 7 ml of W5 (5 mM KCl, 154 mM NaCl, 125 mM CaCl_2 , and 2 mM MES pH 5.7) was added carefully.
8. After another centrifugation for 5 min at 100 g with brakes on maximum supernatant was discarded.
9. Afterwards 7 ml of W5 solution was added to purified protoplasts and was centrifuged for 2 min at 100 g with maximum brakes on.
10. Finally, supernatant was discarded and 1 ml of W5 solution was added to the purified protoplasts before estimating protoplasts yield using a haemocytometer.

PEG mediated protoplasts transformation

1. The purified protoplasts were re-suspended in appropriate amount of MMG solution with a desired concentration of 2×10^6 protoplasts ml^{-1} .
2. Thirty μg plasmid DNA was added to a 2 ml Eppendorf tube. Afterwards 200 μl of the protoplast solution was added to the tube and mixed gently.
3. An equal volume (230 μl) of the PEG solution was then added and mixed thoroughly by gently tapping and inverting the tube several times.

4. The DNA-protoplast-PEG mixture was incubated in the dark for 30 min in room temperature. After 30 min about 2x volume (900 μl) of W5 solution was added to the incubated mixture and mixed well by inverting the tube 2–3 times to stop the transformation process.
5. The mixture was then centrifuged for 5 min at 100 rpm.

Mutation detection in gene edited pooled protoplasts

1. CRISPR/Cas-mediated targeted mutations were detected by PCR using genomic DNA extracted from transfected pooled protoplasts after 72 h of transformation. The primers used in this study are listed in [Supplementary Table 1](#).
2. The targeted gene was then amplified using gene-specific primers and the products were purified by 2% agarose gel electrophoresis in TAE buffer followed by extraction from excised gel bands using an appropriate kit (Qiagen, Hilden, Germany).
3. The purified products were digested with restriction enzymes with recognition sites incorporated in the gRNA mutation target, such that indels would destroy the site and prevent digestion.
4. PCR products that passed this rapid test for editing were then cloned in the TOPO vector system and were sequenced to confirm the mutation.
5. Sequence chromatograph were verified for multiple peaks near 3 bp upstream of the protospacer adjacent motif (PAM) region compared to the non-transformed control.

Results

This study was aimed at establishing optimized protocols for chemical and physical mutagenesis of field pea. The pea seeds were exposed to varying doses of EMS and gamma radiation. Mutagenesis affects various stages of growth and development, especially in the M_1 generation. Along with the deleterious effects of the mutagen on genome integrity, the non-genetic seed components, such as cellular constituents and enzymes are also affected ([Talebi et al., 2012](#); [Kumar et al., 2013](#); [Serrat et al., 2014](#); [Arisha et al., 2015](#)). Here, we quantified several phenotypes at different growth stages of field pea to understand the biological effects of the mutagen and define an optimal dose for mutagenesis.

Assessment of growth phenotypes following chemical mutagenesis

The pea seeds were exposed to EMS doses of 5, 10, 15, 20, 25, and 30 mM. The seed germination rate dropped steadily from ~50% in 5 mM EMS to 37.5 and 10% in 10 mM and 15 mM EMS

doses, and below 5% in the higher concentrations from 20 to 30 mM EMS (Figure 1A). Following a similar pattern, the delay in germination steadily increased with increasing EMS doses. While control treatment (0 mM) showed germination on average in 3 days, the EMS doses of 5, 10, and 15 mM showed germination at 12 days, 17 days, and 23 days, on average, respectively (Figure 1B). The higher doses of EMS (20, 25, and 30 mM) delayed germination to approximately 30 days (Figure 1B). The number of seedlings that survived post-germination followed an inverse pattern to increasing EMS doses. As opposed to 100% survival in control samples, the seeds treated with EMS doses of 5, 10, 15 mM showed 48.5, 36, and 10% seedling survival, respectively, while the higher doses (20, 25, and 30 mM) showed less than 5% seedling survival (Figure 1C). Among the surviving seedlings, we phenotyped the proportion of seedlings with aberrations in the chlorophyll accumulation, referred to here as chlorotic seedlings. The chlorosis in seedlings ranged from non-green patches of varying sizes on leaves to completely albino seedlings. These seedlings likely represent the germplasm with varying degrees of damage to chloroplast function-associated genetic material from chloroplast and nuclear genome. In our study, we observed 3 and 1.5% chlorotic seedlings in 5 and 10 mM EMS treatments, respectively, as compared to no chlorosis in control seedlings. As a measure of growth delay in mutagenized seedlings, we assessed the number of days required to reach the fourth-leaf stage. The control

seedlings reached the fourth-leaf stage in approximately 1 week, while in seedlings treated with 5, 10, and 15 mM EMS the emergence of fourth-leaf occurred in approximately 19, 23, and 29 days, respectively (Figure 1D).

Assessment of growth phenotypes following physical mutagenesis

For physical mutagenesis, the pea seeds were exposed to six doses of gamma radiation: 150, 200, 250, 300, 350, and 400 Gy. The germination rate was only marginally affected at 150 Gy as compared to 93% for the control sample (no radiation treatment; Figure 2A). Radiation treatment of seeds with 200 to 350 Gy showed steady reduction in germination from 71 to 34%, respectively (Figure 2A). The 400 Gy dose showed a similar effect on germination rate as the 350 Gy dose. Gamma irradiation also affected the number of days for germination. While the control seeds germinated on average at 5 days, the seeds irradiated with the lowest dose of 150 Gy germinated on average at 6 days (Figure 2B). The seeds treated with 200 and 250 Gy doses germinated on average at 9 days, delaying the germination by 4 days as compared to the control. Similarly, the doses of 300 and 350 Gy caused the seeds to germinate at 13 and 14 days, and the seeds treated with the strongest dose of 400 Gy showed germination at 16 days on average (Figure 2B).

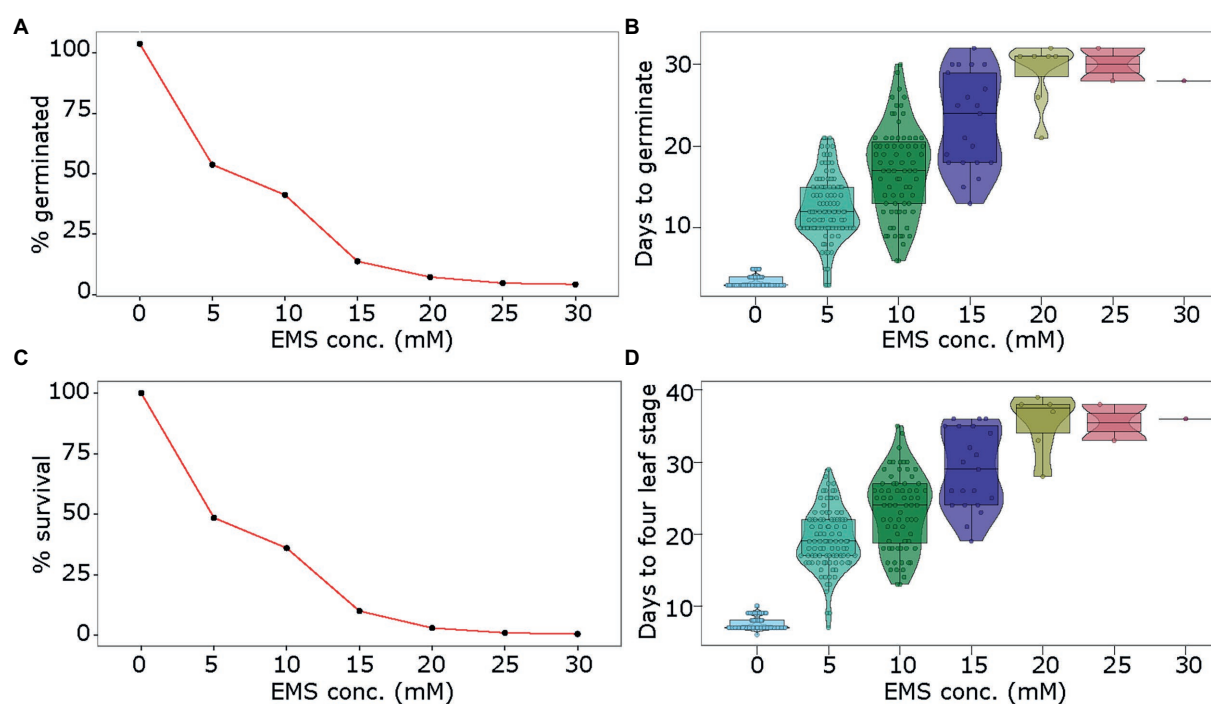


FIGURE 1
Optimization of EMS treatment for mutagenesis of pea seeds. Seeds of pea genotype CDC Amarillo were treated with different concentrations of EMS and were grown in the growth chamber. The germination rate (A), days to germination (B), survival rate (C), and days to fourth-leaf stage (D) were assessed in treated and control samples to identify an optimal dose of EMS for mutagenesis of pea seeds.

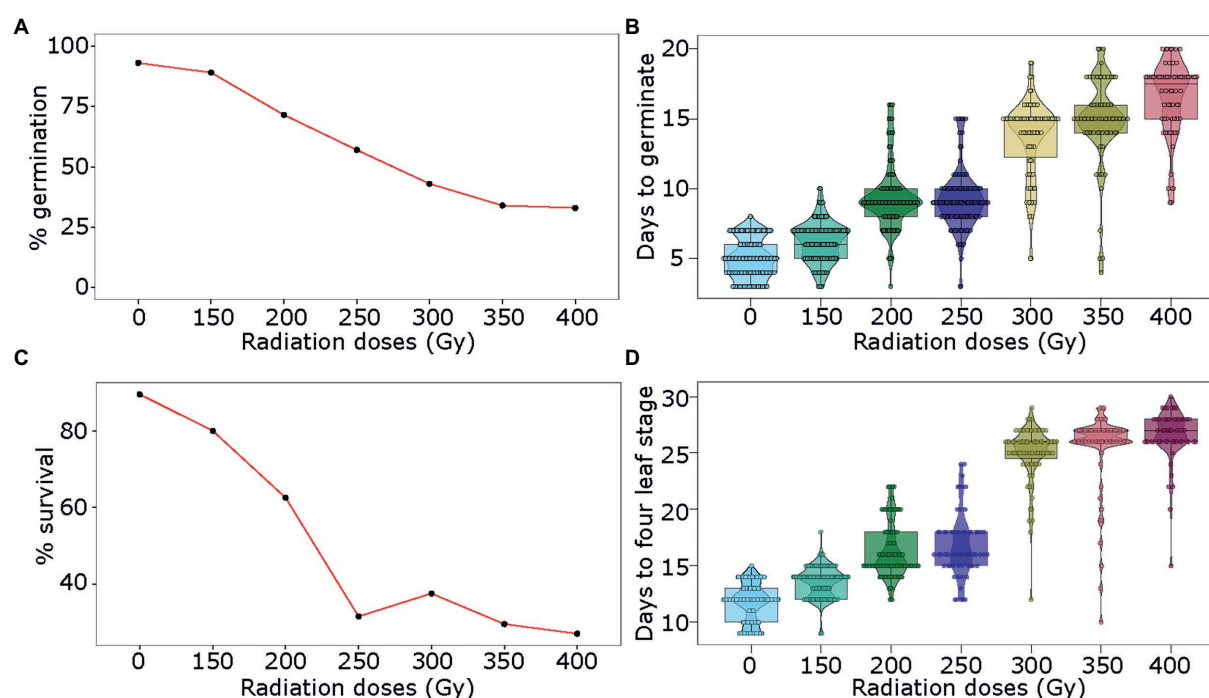


FIGURE 2

Optimization of gamma radiation treatment for mutagenesis of pea seeds. Seeds of pea genotype CDC Amarillo were treated with different doses of gamma radiation and were grown in the growth chamber. The germination rate (A), days to germination (B), survival rate (C), and days to fourth-leaf stage (D) were assessed in treated and control samples to identify an optimal dose of gamma radiation for mutagenesis of pea seed.

Further, the germinated seedlings derived from seeds irradiated with 150 and 200 Gy showed 80 and 62.5% survival rates, respectively (Figure 2C). However, the survival rate dropped between 32 and 27% of the germinated seedlings when the seeds were treated with gamma radiation doses of 250 to 400 Gy (Figure 2C). Of the surviving seedlings, as expected, the strongest radiation dose of 400 Gy showed the highest proportion of chlorotic plants (~14%), whereas lower doses from 150 to 350 Gy showed 2.5–7.5% chlorotic plants. Similar to chemical treatment, early growth retardation was observed for all gamma radiation doses. While the control sample reached the fourth-leaf stage in 12 days after germination, the seedlings from seeds treated with 150, 200, 250, 300, 350, and 400 Gy reached the fourth-leaf stage in 14, 16, 17, 25, 25 and 26 days, respectively (Figure 2D).

Assessment of optimal doses for physical and chemical mutagenesis

The optimal dose for mutagenesis is designated as the dose at which 50% lethality of seeds (LD_{50}) occurs. For chemical mutagenesis of pea seeds, the lowest EMS concentration of 5 mM caused approximately 50% lethality and hence it was chosen as the optimal dose. At this optimal dose, 99.5% of the plants produced seeds. For physical mutagenesis, the radiation dose of 250 Gy caused approximately 50% lethality, however,

only 30% of seedlings survived. On the other hand, the dose of 200 Gy showed 30% seed lethality, with 60% seedling survival (of the initial seeds) rate. Both of these doses showed comparable delays in germination and days to reach the four-leaves stage. Therefore, we selected an intermediate dose of 225 Gy as the optimal dose for physical mutagenesis in field pea. Figure 3 shows comparative phenotypes of pea plants treated with the optimal doses of EMS (upper right panel) and gamma-irradiation (lower right panel) as compared to their respective controls (left panels).

CRISPR/Cas-based editing of LOX gene in pea protoplasts

In this work we first established an efficient pea mesophyll protoplasts isolation and Poly Ethylene Glycol (PEG) mediated transformation system using green fluorescent protein (GFP) as reporter. We used the website² to select the gRNA sequences from the first few exons and introns of the LOX gene. From the list generated, we chose the gRNA sequences with the highest CRISPRater scores which were located in exons and optimized the method for the first time as the evidence of successful CRISPR/Cas-mediated gene editing in pea protoplasts. The

² <https://crispr.cos.uni-heidelberg.de>

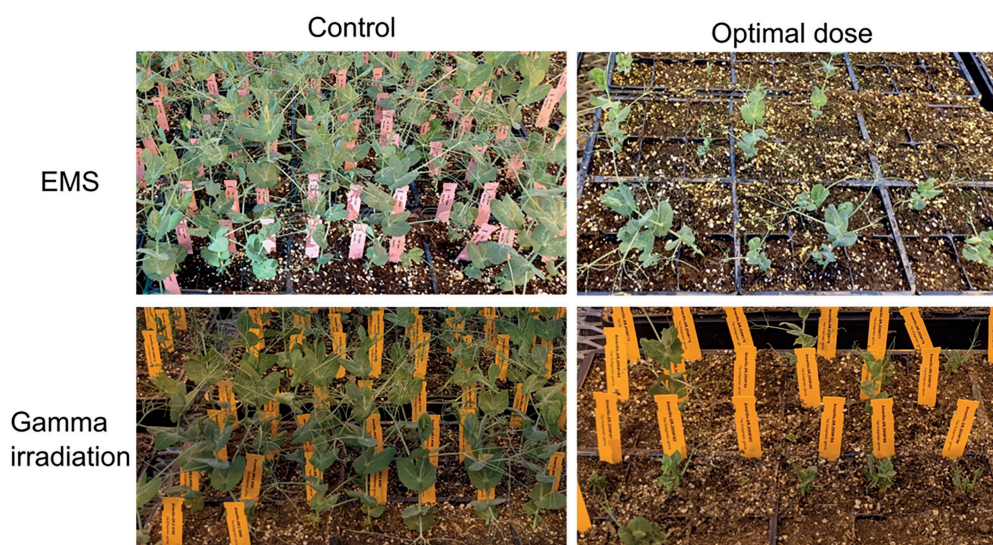


FIGURE 3

Comparative growth phenotypes of plants derived from pea seeds treated with optimized doses of mutagen and their respective controls. The upper panel shows growth phenotype of control (0 mm EMS, upper left panel) and optimal EMS dose (5 mm, upper right panel) treated M_0 pea plants. The lower panel shows growth phenotype of control (untreated, lower left panel) and optimal gamma irradiation dose (250 Gy, lower right panel) treated M_0 pea plants.

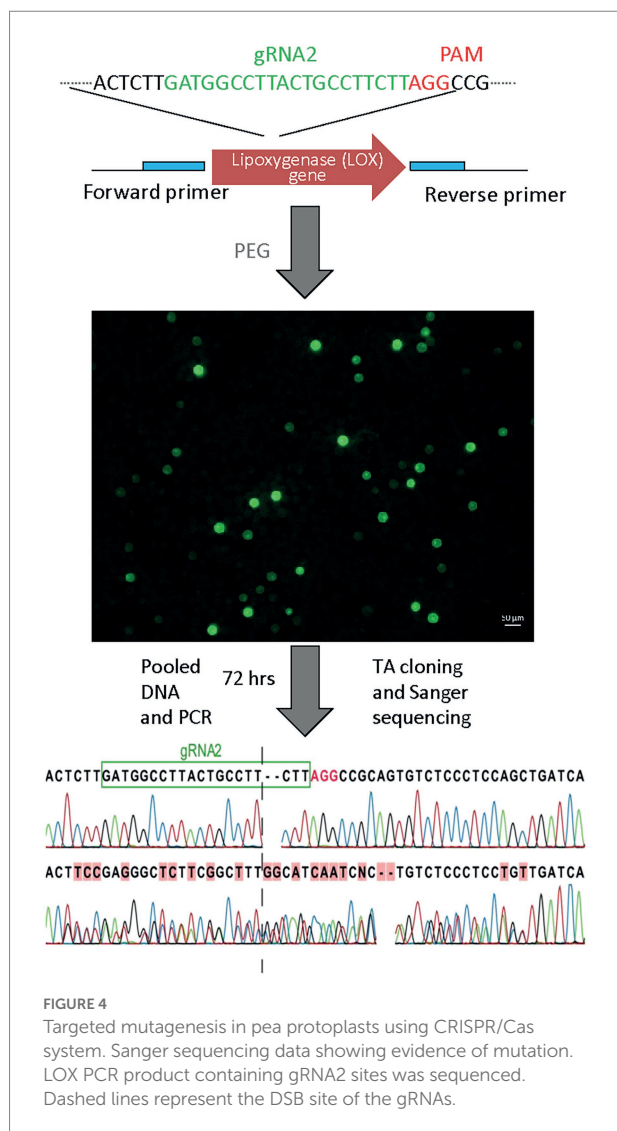
transformation efficiency ($45\% \pm 1.5$) of pea mesophyll protoplasts using the described protocol was quantified by counting the percentage recovery of GFP reporter gene which was co-transformed and served as a control (Figure 4). The LOX PCR product from pooled DNA was gel purified and sequenced using Sanger sequencing (Figure 4). This study demonstrates the value and feasibility of gene editing in pea laying a technical foundation for future trait discovery and improvement of pea and other pulse crops.

Discussion

Field pea is a leguminous break crop and an attractive option for sustainable agricultural practices (Tulbek et al., 2017). It is also a nutrient-rich food crop, and an alternate source of protein, starch, fiber, texturizer, and emulsifier for the exponentially growing plant-based industries (Tulbek et al., 2017). However, the use of field pea as a protein source requires reduction in anti-nutritional compounds, while the field pea-based processing industries require varieties with higher amounts of easily extractable protein, starch, and fiber to increase their profit margins. The establishment of field pea mutant and TILLING resources is an essential prerequisite for the rapid development of such varieties.

In this work, we developed optimized protocols for chemical and physical mutagenesis of pea. The procedures described here could be upscaled to generate a large-scale mutagenized population for pea improvement. We have identified 18 h of incubation in 5 mm EMS to be the optimal dose for chemical

mutagenesis of field pea seeds. This concentration is lower than previous publications which optimized 20 mm EMS for the construction of the green pea TILLING platform and phenotyping database (Dalmais et al., 2008), or 0.3% EMS (corresponding to ~24 mm EMS) for mutational breeding purposes (Sharma et al., 2009). In our study, 20–25 mm EMS treatment led to ~95% seed lethality, while Sharma et al. (2009), reported 53% lethality at 24 mm EMS treatment. Interestingly, Sharma et al. (2009), reported 1.93–3.45% chlorotic individuals in the M_1 generation of 0.2–0.3% EMS (corresponding to 16–24 mm EMS) treated seeds, which is in the same range as 1.5–3% chlorotic individuals observed at 5–10 mm EMS treatment in our study. The variation in results could be attributed to pea varietal differences and the variable times of incubation in different studies. In this study, 18 h of incubation in EMS solutions were employed, while Dalmais et al. (2008) and Sharma et al. (2009) used 15 h and 8 h of incubation in EMS solutions, respectively. Additionally, Dalmais et al. (2008) and Sharma et al. (2009) used garden pea varieties, while we used a field pea cultivar in this study. Cultivar-specific differential responses to mutagens have also been reported previously (Çiftçi et al., 2006; Sharma et al., 2009), which highlights the importance of mutagen dose optimization for different plant cultivars. Interestingly, a previous report of gamma irradiation-mediated mutagenesis of green pea cultivars (Çiftçi et al., 2006) shows similar results as this study. Çiftçi et al. (2006) observed 63–80% and 73–90% germination in different green pea cultivars irradiated at 140 and 180 Gy, respectively, while we observed 89 and 71% germination at similar doses of 150 and 200 Gy, respectively.



Following the optimization of the mutagen doses, the decision of the number of seeds to be used for mutagenesis depends on the aim of the study and the available infrastructure. For building a mutant resource, a large number of seeds need to be mutagenized to ensure multiple mutations per gene or per unit genomic region. For example, the probability of identifying a mutation at a particular G:C base pair in the green pea TILLING database generated from ~4,700 M_1 plants was calculated to be 0.06% (Dalmais et al., 2008). If this resource was generated from 50,000 mutant lines, the probability of identifying mutation at any G: C base pair would have increased to 52% (Dalmais et al., 2008).

Pea belongs to a recalcitrant group of plants in terms of regeneration. If the gene edited protoplasts can be regenerated into plants, protoplasts system could be used to target genes through transient transformation of CRISPR/Cas reagents into protoplasts. By selecting suitable target sites and achieving high editing efficiency, the target gene could be precisely mutated in all the cells of the regenerants if the gene editing event occurs

before cell division. In such scenario, the mutation could be successfully transmitted to the offspring without chromosomal insertion of gene editing components. Nonetheless, the protoplast based gene editing system described here will be a valuable tool for functional genomics research in pea.

Conclusion

We present step-by-step guide of conducting random and directed mutagenesis in commercially important legume crop, field peas. The optimized protocols for pea mutagenesis described here would be useful to create large-scale mutagenized populations to support gene discovery and crop improvement in field pea. The general principles and mutagenesis methods described here also provide a foundation for similar studies in other leguminous crops.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

SK conceived and designed the study. PP performed the chemical and physical mutagenesis experiments. PB performed the gene editing experiments. PP and PB prepared the first draft of the manuscript. SK edited and finalized the manuscript. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the Aquatic and Crop Resource Development Centre as part of its contribution to the Sustainable Protein Production Program of National Research Council Canada.

Acknowledgments

The authors acknowledge the technical support from Ramaswami Sammynaiken and Eli Wiens at the Saskatchewan Structural Sciences Centre, University of Saskatchewan, Saskatoon, SK, Canada. Authors also thank Tom Warkentin for providing germplasm (CDC Meadow), and Connor Lorne Hodgins and Dae-Kyun Ro for help with developing gene editing constructs.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.995542/full#supplementary-material>

References

- Al Bari, M. A., Zheng, P., Viera, I., Worrall, H., Szwiec, S., Ma, Y., et al. (2021). Harnessing genetic diversity in the USDA pea Germplasm collection Through genomic prediction. *Front. Genet.* 12:7754. doi: 10.3389/fgene.2021.707754
- Arisha, M. H., Shah, S. N. M., Gong, Z.-H., Jing, H., Li, C., and Zhang, H.-X. (2015). Ethyl methane sulfonate induced mutations in M2 generation and physiological variations in M1 generation of peppers (*Capsicum annuum* L.). *Front. Plant Sci.* 6:399. doi: 10.3389/fpls.2015.00399
- Bhowmik, P., Konkin, D., Polowick, P., Hodgins, C. L., Subedi, M., Xiang, D., et al. (2021). CRISPR/Cas9 gene editing in legume crops: opportunities and challenges. *Legume Science* 3:e96. doi: 10.1002/leg3.96
- Burstin, J., Salloigne, 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
- Chen, L., Huang, L., Min, D., Phillips, A., Wang, S., Madgwick, P. J., et al. (2012). Development and characterization of a new TILLING population of common bread wheat (*Triticum aestivum* L.). *PLoS One* 7:e41570. doi: 10.1371/journal.pone.0041570
- Çiftçi, C., Türkan, A., and Khawar, K. M. (2006). Use of gamma rays to induce mutations in four pea (*Pisum sativum* L.) cultivars. *Turk. J. Biol.* 1, 29–37.
- Clemente, A., Arques, M. C., Dalmais, M., Signor, C. L., Chinoy, C., Olias, R., et al. (2015). Eliminating anti-nutritional plant food proteins: The case of seed protease inhibitors in pea. *PLoS One* 10:e0134634. doi: 10.1371/journal.pone.0134634
- Crosta, M., Nazzari, N., Ferrari, B., Pecetti, L., Russi, L., Romani, M., et al. (2022). Pea grain protein content Across Italian environments: genetic relationship With grain yield, and opportunities for genome-enabled selection for protein yield. *Front. Plant Sci.* 12:8713. doi: 10.3389/fpls.2021.718713
- Dalmais, M., Schmidt, J., Le Signor, C., Moussy, F., Burstin, J., Savoie, V., et al. (2008). UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol.* 9:R43. doi: 10.1186/gb-2008-9-2-r43
- Davey, M. R., Anthony, P., Power, J. B., and Lowe, K. C. (2005). Plant protoplasts: status and biotechnological perspectives. *Biotechnol. Adv.* 23, 131–171. doi: 10.1016/j.biotechadv.2004.09.008
- Hernández-Muñoz, S., Pedraza-Santos, M. E., López, P. A., Gómez-Sanabria, J. M., and Morales-García, J. L. (2019). Mutagenesis in the improvement of ornamental plants. *Revista Chapingo Serie Horticultura* 25, 151–167. doi: 10.5154/rchsh.2018.12.022
- 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
- Jing, R., Ambrose, M. A., Knox, M. R., Smykal, P., Hybl, M., Ramos, A., et al. (2012). Genetic diversity in European *Pisum* germplasm collections. *Theor. Appl. Genet.* 125, 367–380. doi: 10.1007/s00122-012-1839-1
- Joseph, R., Yeoh, H.-H., and Loh, C.-S. (2004). Induced mutations in cassava using somatic embryos and the identification of mutant plants with altered starch yield and composition. *Plant Cell Rep.* 23, 91–98. doi: 10.1007/s00299-004-0798-7
- Kumar, A. P., Boualem, A., Bhattacharya, A., Parikh, S., Desai, N., Zambelli, A., et al. (2013). SMART – sunflower mutant population And reverse genetic tool for crop improvement. *BMC Plant Biol.* 13:38. doi: 10.1186/1471-2229-13-38
- Li, G., Liu, R., Xu, R., Varshney, R. K., Ding, H., Li, M., et al. (2022). Development of an agrobacterium-mediated CRISPR/Cas9 system in pea (*Pisum sativum* L.). *Crop J.* doi: 10.1016/j.cj.2022.04.011 (in press).
- Nasti, R. A., and Voytas, D. F. (2021). Attaining the promise of plant gene editing at scale. *Proc. Natl. Acad. Sci.* 118:e2004846117. doi: 10.1073/pnas.2004846117
- Oladoso, Y., Rafii, M. Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H. A., et al. (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnol. Biotechnol. Equip.* 30, 1–16. doi: 10.1080/13102818.2015.1087333
- Ookawa, T., Inoue, K., Matsuoka, M., Ebitani, T., Takarada, T., Yamamoto, T., et al. (2014). Increased lodging resistance in long-culm, low-lignin *gh2* rice for improved feed and bioenergy production. *Sci. Rep.* 4:6567. doi: 10.1038/srep06567
- Pandey, A. K., Rubiales, D., Wang, Y., Fang, P., Sun, T., Liu, N., et al. (2021). Omics resources and omics-enabled approaches for achieving high productivity and improved quality in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 134, 755–776. doi: 10.1007/s00122-020-03751-5
- Perera, D., Barnes, D. J., Baldwin, B. S., and Reichert, N. A. (2015). Mutagenesis of in vitro cultures of *Miscanthus × giganteus* cultivar freedom and detecting polymorphisms of regenerated plants using ISSR markers. *Ind. Crop. Prod.* 65, 110–116. doi: 10.1016/j.indcrop.2014.12.005
- Ranjbar Sistani, N., Kaul, H.-P., Desalegn, G., and Wienkoop, S. (2017). Rhizobium impacts on seed productivity, quality, and protection of *Pisum sativum* upon disease stress caused by *Didymella pinodes*: phenotypic, proteomic, and Metabolomic traits. *Front. Plant Sci.* 8:1961. doi: 10.3389/fpls.2017.01961
- Rawal, V., and Navarro, D. K., (eds.) (2019). *The Global Economy of Pulses*. Rome, FAO.
- Schreiber, M., Barakate, A., Uzrek, N., Macaulay, M., Sourdille, A., Morris, J., et al. (2019). A highly mutagenised barley (cv. Golden promise) TILLING population coupled with strategies for screening-by-sequencing. *Plant Methods* 15:99. doi: 10.1186/s13007-019-0486-9
- Serrat, X., Esteban, R., Guibourt, N., Moyssset, L., Nogués, S., and Lalanne, E. (2014). EMS mutagenesis in mature seed-derived rice calli as a new method for rapidly obtaining TILLING mutant populations. *Plant Methods* 10:5. doi: 10.1186/1746-4811-10-5
- Sharma, A., Plaha, P., Rathour, R., Katoch, V., Singh, Y., and Khalsa, G. S. (2009). Induced mutagenesis for improvement of garden pea. *Intern. J. Veg. Sci.* 16, 60–72. doi: 10.1080/19315260903195634
- Shirasawa, K., Sasaki, K., Hirakawa, H., and Isobe, S. (2021). Genomic region associated with pod color variation in pea (*Pisum sativum*). *G3 (Bethesda, Md.)* 11:081. doi: 10.1093/g3journal/jkab081
- Sikora, P., Chawade, A., Larsson, M., Olsson, J., and Olsson, O. (2012). Mutagenesis as a tool in plant genetics, functional genomics, and breeding. *Intern. J. Plant Genom.* 2011:e314829. doi: 10.1155/2011/314829
- Sim, S. Y. J., Sriv, A., Chiang, J. H., and Henry, C. J. (2021). Plant proteins for future foods: A roadmap. *Foods* 10:1967. doi: 10.3390/foods10081967
- Sinha, A., Wetten, A. C., and Caligari, P. D. S. (2003). Effect of biotic factors on the isolation of *Lupinus albus* protoplasts. *Aust. J. Bot.* 51, 103–109. doi: 10.1071/BT01104
- Stagnari, F., Maggio, A., Galieni, A., and Pisante, M. (2017). Multiple benefits of legumes for agriculture sustainability: an overview. *Chem. Biol. Technol. Agric.* 4:2. doi: 10.1186/s40538-016-0085-1
- Talebi, A. B., Talebi, A. B., and Shahrokhifard, B. (2012). Ethyl methane Sulphonate (EMS) induced mutagenesis in Malaysian Rice (cv. MR219) for lethal dose determination. *Am. J. Plant Sci.* 3, 1661–1665. doi: 10.4236/ajps.2012.312202
- Tulbek, M. C., Lam, R. S. H., Wang, Y. C., Asavajaru, P., and Lam, A. (2017). “Chapter 9 - pea: A sustainable vegetable protein crop,” in *Sustainable Protein Sources*. eds. S. R. Nadathur, J. P. D. Wanasundara and L. Scanlin (San Diego: Academic Press), 145–164.

Wang, X., Yu, R., and Li, J. (2021). Using genetic engineering techniques to develop Banana cultivars With Fusarium wilt resistance and ideal plant architecture. *Front. Plant Sci.* 11:7528. doi: 10.3389/fpls.2020.617528

Warkentin, T. D., Vandenberg, A., Tar'an, B., Banniza, S., Arganosa, G., Barlow, B., et al. (2014). CDC Amarillo yellow field pea. *Can. J. Plant Sci.* 94, 1539–1541. doi: 10.4141/cjps-2014-200

Warkentin, T., Vandenberg, A., Taran, B., Banniza, S., Barlow, B., and Ife, S. (2007). CDC meadow field pea. *Can. J. Plant Sci.* 87, 909–910. doi: 10.4141/CJPS07038

Yang, C., Mulligan, B. J., and Wilson, Z. A. (2004). Molecular genetic analysis of pollen irradiation mutagenesis in *Arabidopsis*. *New Phytol.* 164, 279–288. doi: 10.1111/j.1469-8137.2004.01182.x

Zheng, Y., Li, S., Huang, J., Fu, H., Zhou, L., Furusawa, Y., et al. (2020). Mutagenic effect of three ion beams on Rice and identification of heritable mutations by whole genome sequencing. *Plan. Theory* 9:551. doi: 10.3390/plants9050551

Zhu, Y., Mang, H., Sun, Q., Qian, J., Hipps, A., and Hua, J. (2012). Gene discovery using mutagen-induced polymorphisms and deep sequencing: application to plant disease resistance. *Genetics* 192, 139–146. doi: 10.1534/genetics.112.141986



OPEN ACCESS

EDITED BY

Chuanzhi Zhao,
Shandong Academy of Agricultural
Sciences, China

REVIEWED BY

Ling Jiang,
Chinese Academy of Agricultural
Sciences (CAAS), China
Debatosh Das,
University of Missouri, United States

*CORRESPONDENCE

Changsheng Li
csli@ahau.edu.cn

†These authors have contributed
equally to this work

SPECIALTY SECTION

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

RECEIVED 15 June 2022

ACCEPTED 15 August 2022

PUBLISHED 15 September 2022

CITATION

Xiang X, Hu B, Pu Z, Wang L, Leustek T
and Li C (2022) Co-overexpression
of *AtSAT1* and *EcPAPR* improves seed
nutritional value in maize.
Front. Plant Sci. 13:969763.
doi: 10.3389/fpls.2022.969763

COPYRIGHT

© 2022 Xiang, Hu, Pu, Wang, Leustek
and Li. 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(s)
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.

Co-overexpression of *AtSAT1* and *EcPAPR* improves seed nutritional value in maize

Xiaoli Xiang^{1,2†}, Binhua Hu^{1†}, Zhigang Pu¹, Lanying Wang¹,
Thomas Leustek³ and Changsheng Li^{2*}

¹Institute of Biotechnology and Nuclear Technology, Sichuan Academy of Agricultural Sciences, Chengdu, China, ²The National Engineering Laboratory of Crop Stress Resistance Breeding, Anhui Agricultural University, Hefei, China, ³Department of Plant Biology, Rutgers University, New Brunswick, NJ, United States

Maize seeds synthesize insufficient levels of the essential amino acid methionine (Met) to support animal and livestock growth. *Serine acetyltransferase1* (*SAT1*) and *3'-phosphoadenosine-5'-phosphosulfate reductase* (*PAPR*) are key control points for sulfur assimilation into Cys and Met biosynthesis. Two high-MET maize lines *pRbcS:AtSAT1* and *pRbcS:EcPAPR* were obtained through metabolic engineering recently, and their total Met was increased by 1.4- and 1.57-fold, respectively, compared to the wild type. The highest Met maize line, *pRbcS:AtSAT1-pRbcS:EcPAPR*, was created by stacking the two transgenes, causing total Met to increase 2.24-fold. However, the *pRbcS:AtSAT1-pRbcS:EcPAPR* plants displayed progressively severe defects in plant growth, including early senescence, stunting, and dwarfing, indicating that excessive sulfur assimilation has an adverse effect on plant development. To explore the mechanism of correlation between Met biosynthesis in maize leaves and storage proteins in developing endosperm, the transcriptomes of the sixth leaf at stage V9 and 18 DAP endosperm of *pRbcS:AtSAT1*, *pRbcS:AtSAT1-pRbcS:EcPAPR*, and the null segregants were quantified and analyzed. In *pRbcS:AtSAT1-pRbcS:EcPAPR*, 3274 genes in leaves (1505 up- and 1769 downregulated) and 679 genes in the endosperm (327 up- and 352 downregulated) were differentially expressed. Gene ontology (GO) and KEGG (Kyoto encyclopedia of genes and genomes) analyses revealed that many genes were associated with Met homeostasis, including transcription factors and genes involved in cysteine and Met metabolism, glutathione metabolism, plant hormone signal transduction, and oxidation–reduction. The data from gene network analysis demonstrated

that two genes, serine/threonine-protein kinase (CCR3) and heat shock 70 kDa protein (HSP), were localized in the core of the leaves and endosperm regulation networks, respectively. The results of this study provide insights into the diverse mechanisms that underlie the ideal establishment of enhanced Met levels in maize seeds.

KEYWORDS

methionine, serine acetyltransferase1, 3'-phosphoadenosine-5'-phosphosulfate reductase, transcriptome profiling analysis, protein-protein interaction

Introduction

Maize, like many other cereal crops, is one of the main nutritional resources for humans and livestock but lacks the essential amino acid methionine (Met). Met is a sulfur (S)-containing amino acid produced in three enzymatic steps from cysteine (Cys; Takahashi et al., 2011). Met is mainly accumulated in endosperm in Met-rich zein proteins, i.e., 18- and 10-kDa δ -zein and 15-kDa β -zein. The 18- and 10-kDa δ -zein make up less than 5% of the total zeins (Wu et al., 2009). The 15-kDa β -zein comprises 5–10% of total zeins (Thompson and Larkins, 1994). Met deficiency in the kernel is attributed to sulfur reduction in the leaves as a source and a low content of high-MET-containing endogenous proteins in seeds as a sink (Bagga et al., 2005; Wu et al., 2012).

Many approaches have been used to improve the nutritional quality of modern corn. Both the overexpression of 10-kDa δ -zein or knockdown of 22-kDa δ -zein significantly increase Met content but at the expense of other zein proteins and reduced Cys (Lai and Messing, 2002; Wu et al., 2012), indicating that the availability of S-amino acids limits the total S-amino acids in zeins. Serine acetyltransferase (SAT) is an enzyme that catalyzes serine to form O-acetylserine (OAS; Tabe et al., 2010). 3'-phosphoadenosine-5'-phosphosulfate reductase (PAPR) uses 3'-phosphoadenosine-5'-adenylylsulfate (PAPS) as the substrate and thioredoxins as the electron donor to form sulfite (Martin et al., 2005). 5'-adenylylsulfate (APS) reductase and PAPR constitute the overexpression line (Martin et al., 2005). Evidence from enzymological and physiological studies suggests that both reactions are key regulatory metabolic steps in the control of Cys synthesis. Assimilation into Cys occurs when sulfite reacts with OAS catalyzed by OAS thiolase. In our previous study, a metabolic engineering approach leading to high MET with minimal or no perturbation to the plant phenotype was conducted by leaves specifically overexpressing these two key enzymes, SAT and PAPR. Met is increased 1.40- and 1.57-fold in maize seeds of *pRbcS:AtSAT1* (Xiang et al., 2017) and

pRbcS:EcPAPR (Planta et al., 2017), respectively. In addition, 18- and 10-kDa δ -zein and 15-kDa β -zein protein in those lines also increased.

To further increase the Met level in maize kernels, the *pRbcS:AtSAT1* and *pRbcS:EcPAPR* lines were combined through crossbreeding to create a co-overexpression line. The 15-kDa β -zein and 18- and 10-kDa δ -zein content were about 1.5-fold higher than that of either overexpression line. The total Met levels were also much higher than in the single gene overexpression lines. However, the co-overexpression line displayed abnormal plant phenotypes and decreased yield.

Despite these advances in engineered maize with improved nutritional value, little is known concerning the underlying physiological and molecular mechanisms that cause sulfur overproduction in leaves and high-MET storage in the endosperm, which plays an important role in determining its economic and nutritional value.

RNA sequencing (RNA-Seq) in plants is an accurate, fast, and high-performance technology that is used to measure gene expression patterns, study gene function, and induce gene interactions. To explore the mechanism of Met production in leaf tissue, transport, and storage in maize, transcriptome profiling was performed using Null, *pRbcS:AtSAT1*, and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines. Leaf samples were collected from the sixth leaf at stage V9, at which *AtSAT1* and *EcPAPR* are highly expressed (Planta et al., 2017; Xiang et al., 2018). Endosperm tissues were collected 18 days after pollination (DAP) when protein bodies had a highly ordered architecture. The 18- and 10-kDa δ -zein are deposited in the center of the protein body. 15-kDa β -zein is located in the peripheral layer (Lending and Larkins, 1989). Here, the differentially expressed genes (DEGs) involved in MET synthesis, transport, and storage, nutrient reservoir, and plant abiotic stress response in the *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines were summarized, which offers candidate genes for Met transport and storage in seeds for the following research. In addition, the contents of amino acids, zein and non-zein proteins were investigated in this study.

Materials and methods

Genetic materials

The *Arabidopsis thaliana* serine acetyltransferase1 (*AtSAT1*, Gene Bank: BT008309.1) and *Escherichia coli* 3'-phosphoadenosine-5'-phosphosulfate reductase (*EcPAPR*, Gene Bank: NP_417242.1) were both driven by the maize Rubisco small subunit promoter (*pRbcS*, Gene Bank: AH005359.3). They were transferred to the maize B x A HiII hybrid separately and then crossed to "B73" for five generations to create B73 background OE lines. Both have been described separately elsewhere (Planta et al., 2017; Xiang et al., 2017). The obtained OE lines were referred to as *pRbcS:AtSAT1* and *pRbcS:EcPAPR*, respectively. *pRbcS:AtSAT1* and *pRbcS:EcPAPR* were then crossed to produce an *AtSAT1* and *EcPAPR* co-overexpression line, referred to as *pRbcS:AtSAT1-pRbcS:EcPAPR*. The analyzed lines (at least 15 individuals for each line), including *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, *pRbcS:AtSAT1-pRbcS:EcPAPR*, Null segregant and high Met inbred line BSSS53 were grown in a field in Shanghai (N 31 11', E 121 29') in 2016 and 2017.

Amino acid analysis

For leaf samples, the sixth leaf of V9 plants was ground to powder in liquid nitrogen; mature seeds with the embryo removed were ground into fine powder and passed through an 80-mesh sieve. The soluble and total amino acid content of the leaf and kernel samples were analyzed by Beijing Mass Spectrometry Medical Research Co. Ltd. (Beijing, China). The samples were pretreated with performic acid to yield acid-stable derivatives of Cys and Met, cystic acid, and methionine sulfoxide. The samples were acid hydrolyzed to yield the total amino acid content.

Total zein and non-zein extraction

Total zein and non-zein proteins were extracted as previously described (Wu et al., 2012). For non-zein extraction, after removal of the supernatant, as described above, the solid remnant in the tube was washed with 1 mL zein extraction buffer three times to completely remove zein protein. The solid was evacuated for 30 min at 45°C (VAL model, Eppendorf) and then suspended in 1 mL non-zein extraction buffer [2% (vol/vol) 2-mercaptoethanol, 3.75 mM sodium borate (pH10), 0.3% SDS]. The mixture was maintained at 25°C for 2 h and then centrifuged at 15,700 × g (Eppendorf) for 10 min before 100 µl of the supernatant liquid was transferred into a new tube. Finally, 3 µl protein solution and 8 µl protein loading buffer were mixed and analyzed using SDS/PAGE [15% (wt/vol)] to

determine the non-zein accumulation pattern. The gels were stained with Coomassie blue.

Total protein measurement in mature dry seeds

The dry seeds were ground into a powder and passed through an 80-mesh sieve. Crude protein was measured using a modified Kjeldahl method (International, 1997) that determined the total nitrogen in nitrate-containing materials. The sample was digested in sulfuric acid; ammonia was then distilled, and excess acid was titrated. A conversion factor of 6.25 was used for feedstuffs. The total protein content in the dry seeds was calculated.

Ribonucleic acid isolation and RNA-Seq analysis

Total RNA from the sixth leaves (V9 stage) of maize plants was extracted using TRIzol reagent (Invitrogen Inc., Waltham, MA) and purified with the RNeasy Mini Kit after DNase I digestion (Qiagen Inc., Germantown, MD, United States). Total RNA from 18 DAP endosperm was extracted using the same method as described for the leaves, except that endosperm RNA extract buffer was used before the TRIzol reagent. RNA-Seq experiments were conducted with RNA isolated from three biological replicates for the Null, *pRbcS:AtSAT1*, and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines. Three plants from each genotype were mixed to comprise one biological replicate.

RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, United States). The samples with an RNA integrity number (RIN) ≥ 7 were subjected to subsequent analysis. The libraries were constructed using the TruSeq Stranded mRNA LTSample Prep Kit (Illumina, San Diego, CA, United States) according to the manufacturer's instructions. The libraries were then sequenced on the Illumina sequencing platform (HiSeq™ 2500 or other platforms), and 150 bp/125bp paired-end reads were generated. Quality control was performed using FastQC (R). The paired-end reads were aligned to the B73 reference genome and the reference gene model dataset using TopHat/Bowtie2.¹ The reference maize genome (RefGen_v4,²), transcript sequence,³ and gene model annotation files⁴ were

1 <http://ccb.jhu.edu/software/tophat/>

2 http://ftp.ensemblgenomes.org/pub/plants/release-34/fasta/zea_mays/dna/Zea_mays.AGPv4.dna.toplevel.fa.gz

3 http://ftp.ensemblgenomes.org/pub/plants/release-34/fasta/zea_mays/cdna/Zea_mays.AGPv4.cdna.all.fa.gz

4 http://ftp.ensemblgenomes.org/pub/plants/release-34/gff3/zea_mays/Zea_mays.AGPv4.34.gff3.gz

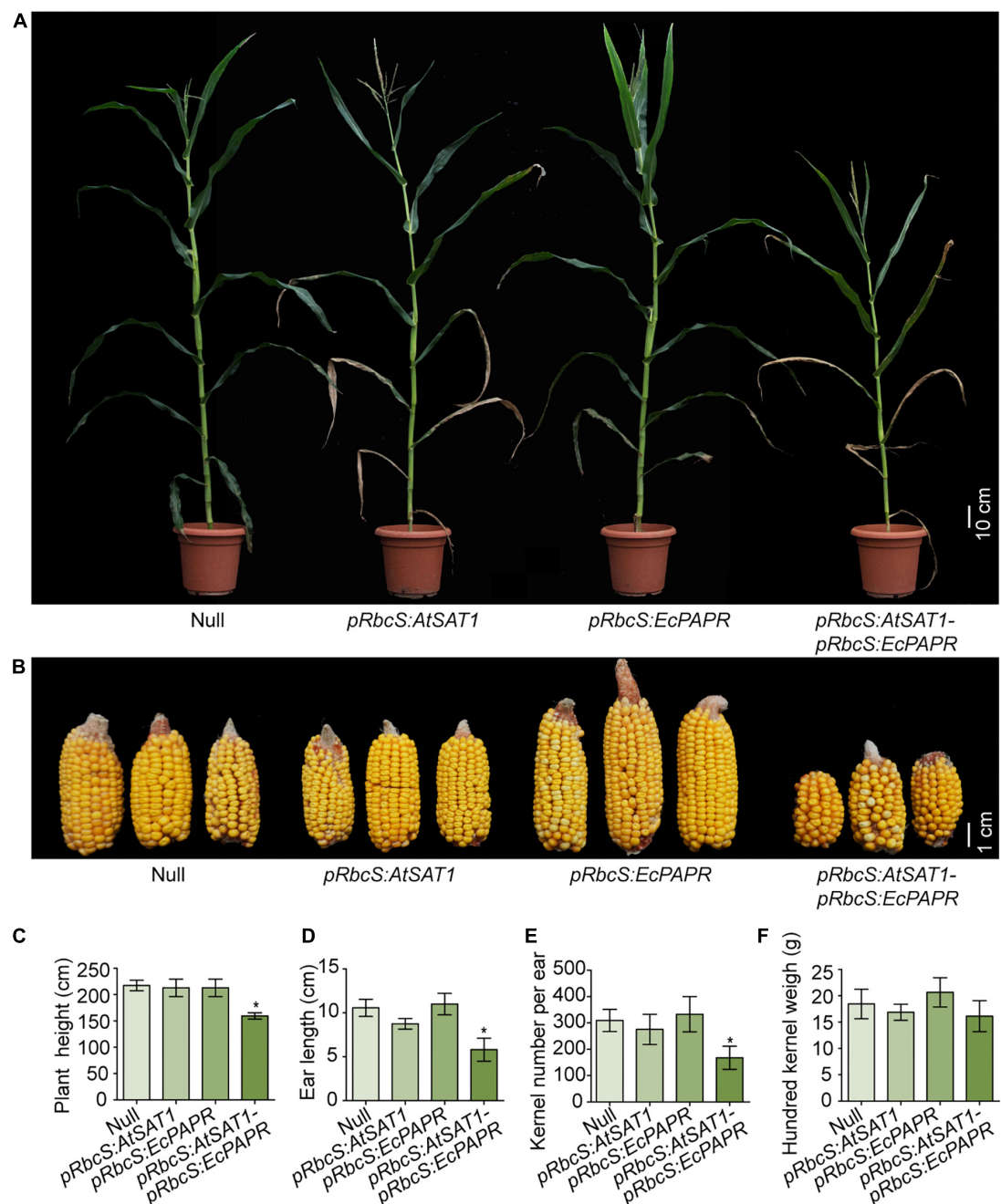


FIGURE 1 Performance of transgenic lines under field conditions. **(A)** Representative photo of flowering plants for each type as labeled in the photograph. **(B)** Representative photos of ears for each type as labeled in the photograph. **(C–F)** Comparison of plant height **(C)**, ear length **(D)**, kernel per ear **(E)**, hundred kernel weight **(F)** between Null, *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, and *pRbcS:AtSAT1-pRbcS:EcPAPR*. The data shown are the means from 15 individual plants \pm SD. Asterisks indicate significant differences from Null (Student's *t*-test, **p*-value < 0.05).

downloaded from Ensemble Genomes.⁵ Gene expression levels in the Null, *pRbcS:AtSAT1*, and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines were determined by normalized FPKM (fragment per

kilobase of transcript (exon model) per million mapped reads) values (Mortazavi et al., 2008; Mizrahi et al., 2010). To determine the variation in expression between three replicates from the Null line and three replicates from *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR*, the absolute difference of the \log_2 -fold change was calculated (*p*-value \leq 0.05). Differentially

⁵ <http://ensemblgenomes.org/>

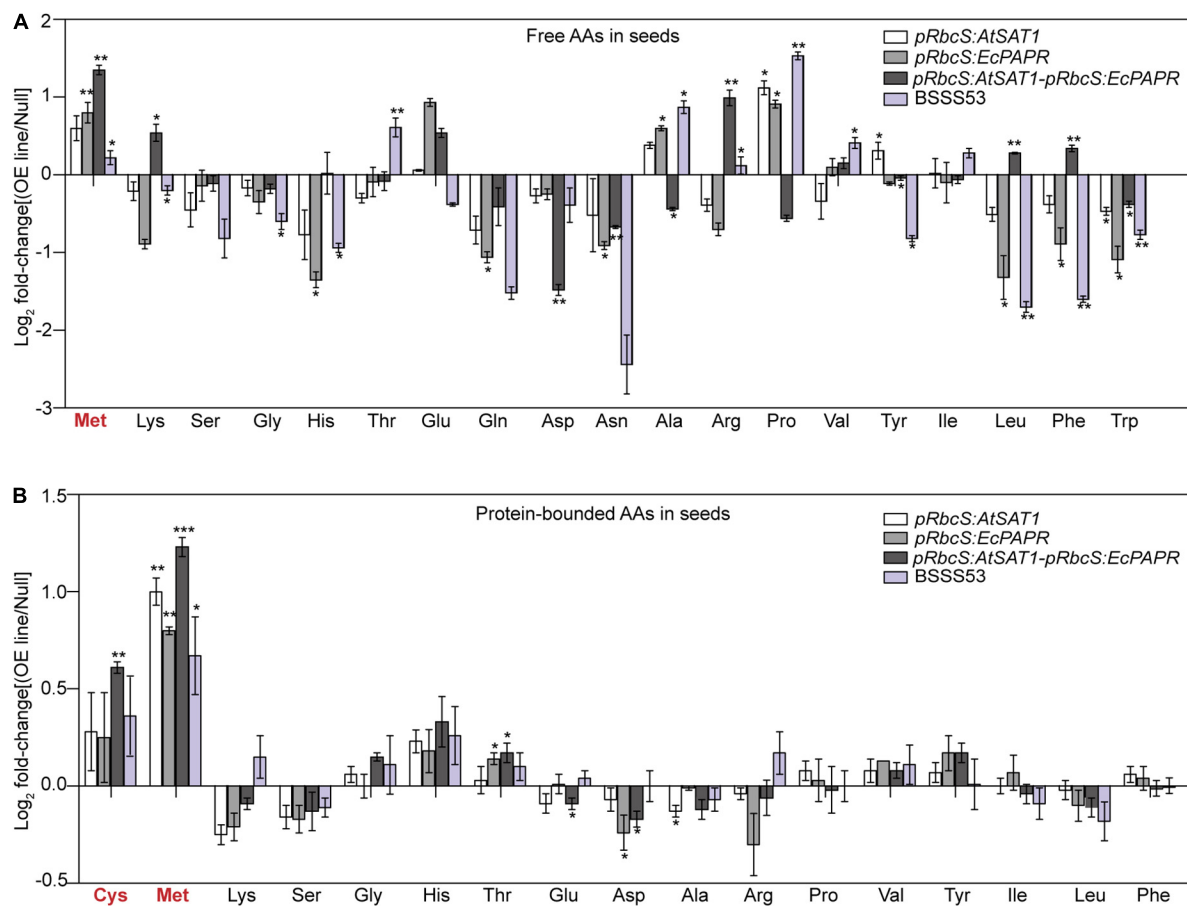


FIGURE 2

Foldchanges in free and protein-bound amino acid levels in transgenic seeds compared with Null. Data were \log_2 -transformed and plotted in the bar graph. (A) Free amino acid levels. (B) Protein-bound amino acid levels. Bars to the left and right indicate a reduction and an increase, respectively, in the amino acid content of the *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, *pRbcS:AtSAT1-pRbcS:EcPAPR*, and BSSS53 plants relative to Null. Student's *t*-test at p -value < 0.05 was used to determine the statistical significance of differences between the transgenic *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, *pRbcS:AtSAT1-pRbcS:EcPAPR*, BSSS53, and non-transgenic Null kernels. Data shown are means \pm SD of three replicates. ** p -value ≤ 0.01 and *** p -value ≤ 0.001 .

expressed genes (DEGs) were identified using a false discovery rate (FDR) threshold.

Gene-level quantification, analysis of differentially expressed genes, cluster analysis, gene ontology, and Kyoto encyclopedia of genes and genomes enrichment

The FPKM value of each gene was calculated using cufflinks (Trapnell et al., 2012), and the read counts of each gene were obtained by htseq-count (Anders et al., 2015). DEGs were identified using the DESeq functions estimateSizeFactors and nbinomTest (Anders, 2012). p -value < 0.05 and fold change > 2 or fold change < 0.5 were set as the thresholds for significant differential expression. Hierarchical cluster analysis of DEGs

was performed to explore gene expression patterns. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) pathway enrichment analyses of DEGs were performed using R based on the hypergeometric distribution (Team, 2014).

Real-time quantitative reverse transcription-PCR

Total RNA extracted as described above was used for reverse transcription with the SuperScript III First Strand Kit (Invitrogen). quantitative reverse transcription-PCR (qRT-PCR) was performed as previously described. The following PCR program was used: 95°C for 120 s, followed by 40 cycles of 95°C for 5 s, 60°C for 30 s, and 95°C for 5 s, 65–95°C melt curve, increments of 0.5°C for 5 s. The gene expression levels relative to the maize *ZmActin* (Zm00001d012277) gene were analyzed

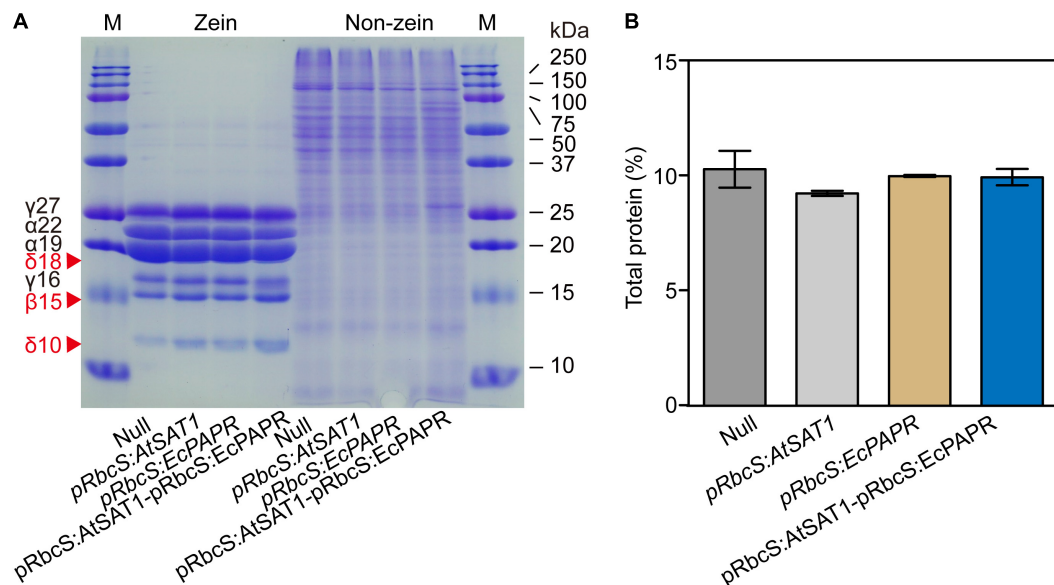


FIGURE 3

Zein and non-zein accumulation pattern in transgenic kernels. (A) Kernels from Null, *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, and *pRbcS:AtSAT1-pRbcS:EcPAPR* were harvested from field plants. The kernels were fully mature, and protein profiles from three different kernels are shown. Protein from 300 μ g dry weight of endosperm sample was loaded in each lane. The mass of each zein is indicated to the right of the figure. (B) Total protein of mature dry seeds.

using the $2^{-\Delta\Delta C_t}$ analysis method. The primers are listed in [Supplementary Table 8](#).

Weighted gene co-expression network analysis and module preservation analysis

Weighted gene co-expression network analysis (WGCNA; Langfelder and Horvath, 2008) based on the differentially expressed (DE) genes package in R was applied to construct a co-expression network using the expression values of 2425 DEGs from *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves and endosperm upregulated DEGs. Clusters were obtained using the molecular complex detection (MCODE) algorithm. Degree centrality analysis of the co-expression network was performed to explore hub genes present in *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves.

The protein interaction data were selected from the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) 9.1 database, and a network was constructed by linking sulfur reduction-related genes with the selected gene signatures using Cytoscape 3.1.0, which is a free software package for visualizing, modeling, and analyzing the integration of biomolecular interaction networks using high-throughput expression data and other molecular states (Shannon et al., 2003).

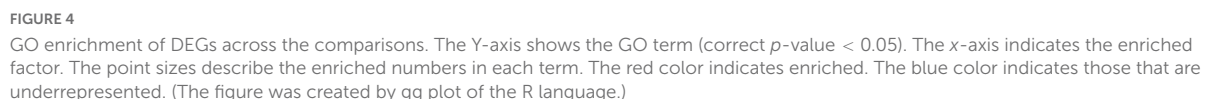
Subsequently, we investigated the substructure of the protein-protein interaction (PPI) network extracted from the

above-constructed network and focused on highly connected nodes, known as clusters, using the MCODE (Bader and Hogue, 2003) clustering algorithm, which includes vertex weighting, complex prediction, and optional post-processing. The core-clustering coefficient was proposed as a metric for sorting the vertices in a graph with respect to their local neighborhood density. The software of the MCODE algorithm was obtained from <http://baderlab.org/Software/MCODE>. The highly interacting nodes in the clusters were identified by parameters keeping K-core = 4, node score cut-off = 0.3, and max depth up to 100.

Results

pRbcS:AtSAT1-pRbcS:EcPAPR co-overexpression line showed abnormal growth due to increasing sulfite assimilation

The *pRbcS:AtSAT1* maize transgenic line was created by overexpressing *AtSAT1* in maize leaves, driven by the leaf-specific bundle-sheath-specific promoter *RbcS1* (Xiang et al., 2017). The *pRbcS:EcPAPR* maize transgenic line was created using the same promoter to drive *EcPAPR* overexpression in maize, as described previously (Planta et al., 2017). The expression of high-MET protein 18- and 10-kDa δ -zein and 15-kDa β -zein was increased in the kernels of both



pRbcS:AtSAT1 has dry leaf tips due to senescence (Xiang et al., 2017). *pRbcS:EcPAPR* shows no phenotypic abnormalities (Planta et al., 2017). We expected that the co-overexpression

of SAT and PAPR would have little or no side effects on plant growth due to the increasing assimilation of sulfite. However, the *pRbcS:AtSAT1-pRbcS:EcPAPR* line showed severely stunted growth, chlorotic leaves, and early senescence at the V5 stage until harvest (Figure 1A). A similar phenotype was obtained by the co-overexpression of *PaAPR* and *EcPAPR* (Martin et al., 2005), but the phenotype described here was more severe than them. The size of the ears and the total kernel number decreased significantly (Figure 1B). In addition, the plant height (Figure 1C), ear length (Figure 1D), and kernel number per ear (Figure 1E) were also significantly decreased when compared with the Null, *pRbcS:AtSAT1*, and *pRbcS:EcPAPR* lines. However, the 100-kernel weight was not significantly different (Figure 1F).

The *pRbcS:AtSAT1-pRbcS:EcPAPR* line had the highest free and protein-bound Met content in leaves and mature seeds

The *pRbcS:AtSAT1-pRbcS:EcPAPR* line inflated pools of free Cys and Met significantly more than *pRbcS:AtSAT1* and *pRbcS:EcPAPR* in the source leaves (Supplementary Figure 1A). Thus, *pRbcS:AtSAT1-pRbcS:EcPAPR* had a larger sulfur amino acid pool than the single gene overexpression lines *pRbcS:AtSAT1* and *pRbcS:EcPAPR*. The former line can supply more resources for seed sinks to synthesize sulfur-containing proteins. The contents of other free amino acids (AAs), such as Lys, Gly, Ile, Leu, Phe, and Trp, were increased in the *pRbcS:AtSAT1-pRbcS:EcPAPR* line. The contents of protein-bound Met and Cys were also significantly increased in *pRbcS:AtSAT1-pRbcS:EcPAPR* (Supplementary Figure 1B). In addition, the content of other AAs, including Lys, Gly, and Ile, showed similar increases among all high-MET lines.

The composition of mature seeds represents the final result of gene expression throughout seed development (Tabé et al., 2010). To assess the effect of SAT1 and PAPR co-overexpression in the source leaves on maize seed composition at the end of maturation, free and protein-bound AAs were detected in hydrolyzed flour from pooled mature dry seed samples of Null, *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, *pRbcS:AtSAT1-pRbcS:EcPAPR*, and BSS53 lines, with three replicates. Maize inbred line BSS53 is a high MET variant (Benner et al., 1989). Both free and protein-bound MET increased in the sink kernel, and the AA profile was changed to form a better nutrition pattern in the *pRbcS:AtSAT1-pRbcS:EcPAPR* line (Figures 2A,B). Free Lys, Arg, Leu, and Phe were increased in the *pRbcS:AtSAT1-pRbcS:EcPAPR* line compared with the Null and all other high-MET lines (Figure 2A). The major changes in the AA profile in the protein-bound AAs were increased Cys and Met, which were significantly higher in the *pRbcS:AtSAT1-pRbcS:EcPAPR*

line (Figure 2B). The *pRbcS:AtSAT1-pRbcS:EcPAPR* line had the highest sulfur-AA content at 1.1% Cys and 4.46% Met.

Kernel nutritional value improved in the *pRbcS:AtSAT1-pRbcS:EcPAPR* line without affecting the seeds' total protein level

In maize (*Zea mays*), the endosperm storage proteins zein and non-zein constitute a major protein component of the seed (Wu and Messing, 2010). Zein and non-zein of mature endosperm were analyzed by SDS-PAGE. Compared to the Null, all OE lines had higher expression of 18- and 10-kDa δ -zein and 15-kDa β -zein (Figure 3A). Moreover, the *pRbcS:AtSAT1-pRbcS:EcPAPR* line had the highest levels of 10-kDa δ -zein and 15-kDa β -zein and showed significantly higher expression levels than the single gene overexpression lines *pRbcS:AtSAT1* and *pRbcS:EcPAPR* (Figure 3A). The expression levels of 22- and 19-kDa γ -zein were slightly decreased in the *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines. The 27- and 16-kDa γ -zein of the *pRbcS:AtSAT1-pRbcS:EcPAPR* line were not increased compared with either *pRbcS:AtSAT1* or *pRbcS:EcPAPR*. The non-zein portions of the Null, *pRbcS:AtSAT1*, and *pRbcS:EcPAPR* lines were not significantly different (Figure 3A). However, the non-zein content of *pRbcS:AtSAT1-pRbcS:EcPAPR* was slightly higher than that of the other three lines (Figure 3A), while the total protein in all lines was not significantly different (Figure 3B).

Leaves of the high-MET line had 10.0-fold more differentially expressed genes than the endosperm

Transcriptomic analysis using the sixth leaves at the V9 stage and 18 DAP endosperm was carried out using Illumina RNA-Seq. About 80% of the raw reads from each sample were mapped to annotated gene-coding regions (Supplementary Table 1). Based on the global FPKM-expressing values (corrected p -value < 0.05), principal component analysis (PCA) showed that the biological replicates were clustered together, indicating experimental consistency. Hierarchical cluster analysis also showed a high correlation within sample replicates (Supplementary Figures 2A,B,E,F).

Compared to the leaf samples from the Null line, *pRbcS:AtSAT1* had 2158 DEGs (1327 up- and 831 downregulated genes; Supplementary Table 2). *pRbcS:AtSAT1-pRbcS:EcPAPR* had 3274 DEGs (1505 up- and 1769 downregulated genes; Supplementary Table 3). There were 1032 specific DEGs in *pRbcS:AtSAT1*, and 2148 DEGs in *pRbcS:AtSAT1-pRbcS:EcPAPR*; they shared 1126 common DEGs (Supplementary Figures 2C,D).

Compared to the endosperm samples from the Null line, *pRbcS:AtSAT1* had 307 DEGs (137 up- and 170 downregulated genes; **Supplementary Table 4**). *pRbcS:AtSAT1-pRbcS:EcPAPR* had 697 DEGs (327 up- and 352 downregulated genes; **Supplementary Table 5**). *pRbcS:AtSAT1* had 188 unique DEGs, and *pRbcS:AtSAT1-pRbcS:EcPAPR* had 460, while they shared 120 DEGs (**Supplementary Figures 2C,D**).

The *pRbcS:AtSAT1-pRbcS:EcPAPR* line contained more DEGs than the *pRbcS:AtSAT1* line, suggesting that the co-overexpression of *AtSAT1* and *EcPAPR* affects more genes than the *AtSAT1*-overexpression line. The number of DEGs in the leaf samples was 10-fold higher than in the endosperm, indicating that the leaf source could be the main limitation of the Met pool in the seed sink.

Gene ontology and Kyoto encyclopedia of genes and genomes analysis revealed differentially expressed genes mainly enriched in the Met and glutathione pathway

To further reveal the functional roles of these DEGs involved in Met biosynthesis and storage. All up- and downregulated DEGs were used to investigate the GO annotations and in the KEGG pathway enrichment analysis (**Supplementary Figures 3, 4, and 5**). We further performed an enrichment analysis of sulfur-related GO terms (**Figure 4**). Upregulated genes enriched in both *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves included “sulfite reductase activity,” “methionine biosynthesis activity,” “glutamate synthase activity,” “cysteine biosynthetic process,” “amino acid transport,” “glutathione transferase activity,” and “sulfate reduction” (**Figure 4**). These indicated that both overexpression lines strengthened the expression pattern of sulfur-related pathway genes. For KEGG pathway enrichment of leaf tissue, many upregulated genes in both the *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines were associated with “sulfur metabolism,” “cysteine and methionine metabolism,” and “glutathione metabolism” (**Supplementary Figure 5**). Up- and downregulated genes enriched in both *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves also included the “oxidation-reduction process,” indicating that the plant oxidation–reduction level changed in many aspects.

For KEGG pathway enrichment of endosperm tissue, upregulated genes in *pRbcS:AtSAT1* included “glutathione metabolism.” Upregulated genes in *pRbcS:AtSAT1-pRbcS:EcPAPR* included “biosynthesis of unsaturated fatty acid” and “fatty acid biosynthesis.” Downregulated genes in *pRbcS:AtSAT1-pRbcS:EcPAPR* included “peroxisome,” “nitrogen metabolism,” “cyanoamino acid metabolism,” “arginine and proline metabolism,” “carbon metabolism,” and “alanine, aspartate and glutamate metabolism” (**Supplementary Figure 5**).

Met biosynthesis genes were upregulated in *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves

The fold change of up- and downregulated DEGs in *pRbcS:AtSAT1-pRbcS:EcPAPR* was twice as high as the related DEGs in *pRbcS:AtSAT1* (**Supplementary Tables 2–5**). Upregulated genes related to sulfur metabolism, glutathione metabolism, amino acid transporters, and metal cation transporters are listed in **Table 1**. Several key regulatory genes involved in the sulfate reduction pathway were significantly upregulated in *pRbcS:AtSAT1-pRbcS:EcPAPR*, including *ATP sulfurylase1* (*APS1*), which adenylates sulfate (SO_4^{2-}) to form 5'-adenylylsulfate (*APS*); *adenosine 5'-phosphosulfate reductase-like1* (*APRL1*) and *adenosine 5'-phosphosulfate reductase-like2* (*APRL2*), both of which catalyze sulfate to form sulfite (SO_3^{2-}); *sulfite reductase 1* (*SiR1*), which catalyzes sulfite to form sulfide (S^{2-}); *serine acetyltransferase2* (*SAT2*), which catalyzes serine to form O-acetylserine (*OAS*); and S^{2-} reacts with *OAS* catalyzed by *OAS* thiol-lyase to form Cys (**Table 1**).

Nutrient reservoir-related gene expression patterns were reformed in the *pRbcS:AtSAT1-pRbcS:EcPAPR* endosperm

pRbcS:AtSAT1-pRbcS:EcPAPR had more up- and downregulated DEGs than *pRbcS:AtSAT1* in 18 DAP endosperm (**Supplementary Tables 4 and 5**). The fold change in DEG expression in *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves almost doubled compared to *pRbcS:AtSAT1*, while the endosperm values of these two high-MET lines were not significantly different. These results indicate that transgenic line *pRbcS:AtSAT1-pRbcS:EcPAPR* has a slightly greater effect on seed sink protein accumulation.

Previous results showed that *pRbcS:AtSAT1* increased MET-rich zein, 18- and 10-kDa δ -zein and 15-kDa β -zein, by increasing mRNA transcription levels (Xiang et al., 2017). The nutrient reservoir activity-related DEGs, including zein genes, were up- or downregulated in *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* transcriptomic analysis (**Figure 5A** and **Table 2**). Most of the increased zeins were high-MET zeins, including *dzs18* (Zm00001d037436), which increased 2.51-fold in *pRbcS:AtSAT1* and 3.33-fold in *pRbcS:AtSAT1-pRbcS:EcPAPR*, and *dzs10* (Zm00001d045937), which was increased 2.00-fold in *pRbcS:AtSAT1* and 3.08-fold in *pRbcS:AtSAT1-pRbcS:EcPAPR*. Zein expression patterns showed similar results as those from the qRT-PCR of the Null, *pRbcS:AtSAT1*, *pRbcS:EcPAPR*, and *pRbcS:AtSAT1-pRbcS:EcPAPR* lines. The expression of *dzs18*, *dzs10*, and *zp15* in the *pRbcS:AtSAT1-pRbcS:EcPAPR* line was significantly improved (**Figure 5B**).

TABLE 1 Partial Up-regulated genes in *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves.

	Gene ID (v3)	<i>pRbcS:AtSAT1</i>		<i>pRbcS:AtSAT1-pRbcS:EcPAPR</i>		Description
		Fold change	P-value	Fold change	P-value	
Up-regulated S metabolism genes	Zm00001d021168	32.2	1.85E-18	79.6	1.46E-14	<i>UGT74F1</i> transfers UDP
	Zm00001d020592	–	–	36.46	0.016549	<i>Glutelin-2</i>
	Zm00001d046226	7.4	3.03E-08	16.5	9.12E-11	<i>mrpa1</i> – multidrug resistance protein associated1
	Zm00001d033981	–	–	2.3	0.0054	<i>aps1</i> – ATP sulfurylase1
	Zm00001d048189	10.0	2.41E-07	–	–	<i>sulfate transporter 1;3 (SULTR1;3)</i>
	Zm00001d021596	–	–	4.4	2.26E-08	<i>apr11</i> – adenosine 5'-phosphosulfate reductase-like1
	Zm00001d006467	2.3	3.76E-03	3.5	7.43E-07	<i>apr12</i> – adenosine 5'-phosphosulfate reductase-like2
	Zm00001d038625	2.8	6.59E-12	5.4	7.61E-15	<i>sir1</i> – sulfite reductase1
Up regulated genes in GSH metabolism and metal transporters	Zm00001d028154	–	–	2.4	0.000149	<i>sat2- sat2</i> – serine acetyltransferase2
	Zm00001d035445	8.2	4.76E-09	11.1	3.20E-11	<i>gsh1</i> – gamma-glutamylcysteine synthetase1
	Zm00001d010950	–	–	3.5	6.85E-07	<i>glutamate</i> – cysteine ligase B, chloroplastic
	Zm00001d043845	2.4	9.76E-09	2.3	1.24E-07	<i>glutamate synthase 1 [NADH]</i> , chloroplastic
	Zm00001d024963	6.5	6.86E-06	11.6	4.60E-42	<i>gst22</i> – glutathione transferase22
	Zm00001d042104	48.01	1.71E-22	102.2	3.97E-11	<i>gst7</i> – glutathione transferase7
	Zm00001d018809	16.32	4.57E-07	37.4	1.36E-65	<i>gst6</i> – glutathione transferase6
	Zm00001d024839	13.22	1.04E-16	34.3	6.55E-32	<i>gst2</i> – glutathione S-transferase2
	Zm00001d042096	10.27	1.19E-14	25.7	3.92E-19	<i>gst21</i> – glutathione S-transferase21
	Zm00001d029696	7.2	3.39E-05	19.0	5.35E-19	<i>gst34</i> – glutathione transferase34
	Zm00001d034356	4.7	5.42E-17	11.7	6.48E-17	<i>gst5</i> – glutathione transferase5
	Zm00001d029704	3.0	2.26E-09	5.5	2.28E-20	<i>gst37</i> – glutathione transferase37
	Zm00001d029708	7.1	8.90E-04	5.0	1.03E-10	<i>gst30</i> – glutathione transferase30
	Zm00001d042225	–	–	4.4	9.79E-08	<i>heavy metal transport/detoxification protein</i>
	Zm00001d010410	–	–	2.4	8.93E-05	<i>metal cation transporter putative expressed</i>
	Zm00001d014569	–	–	2.6	6.03E-07	<i>ctap1</i> -copper-transporting atpase p-type 1
	Zm00001d015829	–	–	2.6	1.51E-06	<i>ctap2</i> -copper-transporting atpase p-type 1
	Zm00001d019228	4.2	6.84E-13	6.0	1.64E-13	<i>metal cation transporter</i>
	Zm00001d003616	–	–	2.41	3.20E-07	<i>glycolipid transporter activity</i>
	Zm00001d036965	10.84	2.76E-45	24.0	3.75E-33	<i>ZT4</i> , Zinc transporter 4, metal cation transporter
	Zm00001d034145	–	–	2.2	6.05E-05	<i>ZPI</i> , Zinc-finger protein 1, salt tolerance protein
	Zm00001d003195	–	–	3.1	0.058149	<i>STO</i> , Salt tolerance protein (STO)
	Zm00001d049954	5.1	3.82E-06	6.6	1.74E-10	<i>MDAR</i> -Monodehydroascorbate reductase
	Zm00001d049265	–	–	7.1	3.62E-09	<i>OMT</i> -O-methyltransferase
Up-regulated genes in MET metabolism and AAs transporters	Zm00001d040697	2.9	2.21E-09	3.4	0.000527	<i>SAMS3</i> S-adenosylmethionine synthetase
	Zm00001d048060	2.1	5.12E-05	2.7	0.032008	<i>HMT</i> -Homocysteine S-methyltransferase 1
	Zm00001d042135	5.8	4.89E-08	11.8	3.48E-12	<i>AATT</i> -amino acid transmembrane transporter
	Zm00001d019225	3.9	4.67E-13	5.0	0.000173	<i>AAT</i> -amino acid transporter
	Zm00001d044533	–	–	2.5	6.29E-08	<i>transmembrane transport</i>
Up-regulated TFs	Zm00001d048681	–	–	836.7	3.69E-29	<i>gras84</i> – GRAS-transcription factor 84
	Zm00001d048682	–	–	438.7	6.25E-27	<i>gras82</i> – GRAS-transcription factor 82
	Zm00001d022442	–	–	6.0	0.104687	<i>bzip58</i> – bZIP-transcription factor 58
	Zm00001d005208	–	–	5.6	8.79E-05	<i>nactf5</i> – NAC-transcription factor 5
	Zm00001d034418	–	–	5.1	4.52E-06	<i>hagt78</i> – GNAT-transcription factor 8
	Zm00001d010399	–	–	4.3	1.68E-15	<i>wrky92</i> – WRKY-transcription factor 92
	Zm00001d043491	12.31	5.86E-09	19.3	5.97E-22	<i>ereb134</i> – AP2-EREBP-transcription factor 134
	Zm00001d039245	2.7	1.41E-07	5.0	7.09E-16	<i>wrky93</i> – WRKY-transcription factor 93
	Zm00001d012527	2.9	1.16E-04	4.9	2.25E-06	<i>nactf23</i> – NAC-transcription factor 23
	Zm00001d006001	2.3	6.18E-07	4.5	6.21E-13	<i>wrky71</i> – WRKY-transcription factor 71
	Zm00001d034601	3.1	9.16E-09	4.3	4.81E-10	<i>nactf49</i> – NAC-transcription factor 49

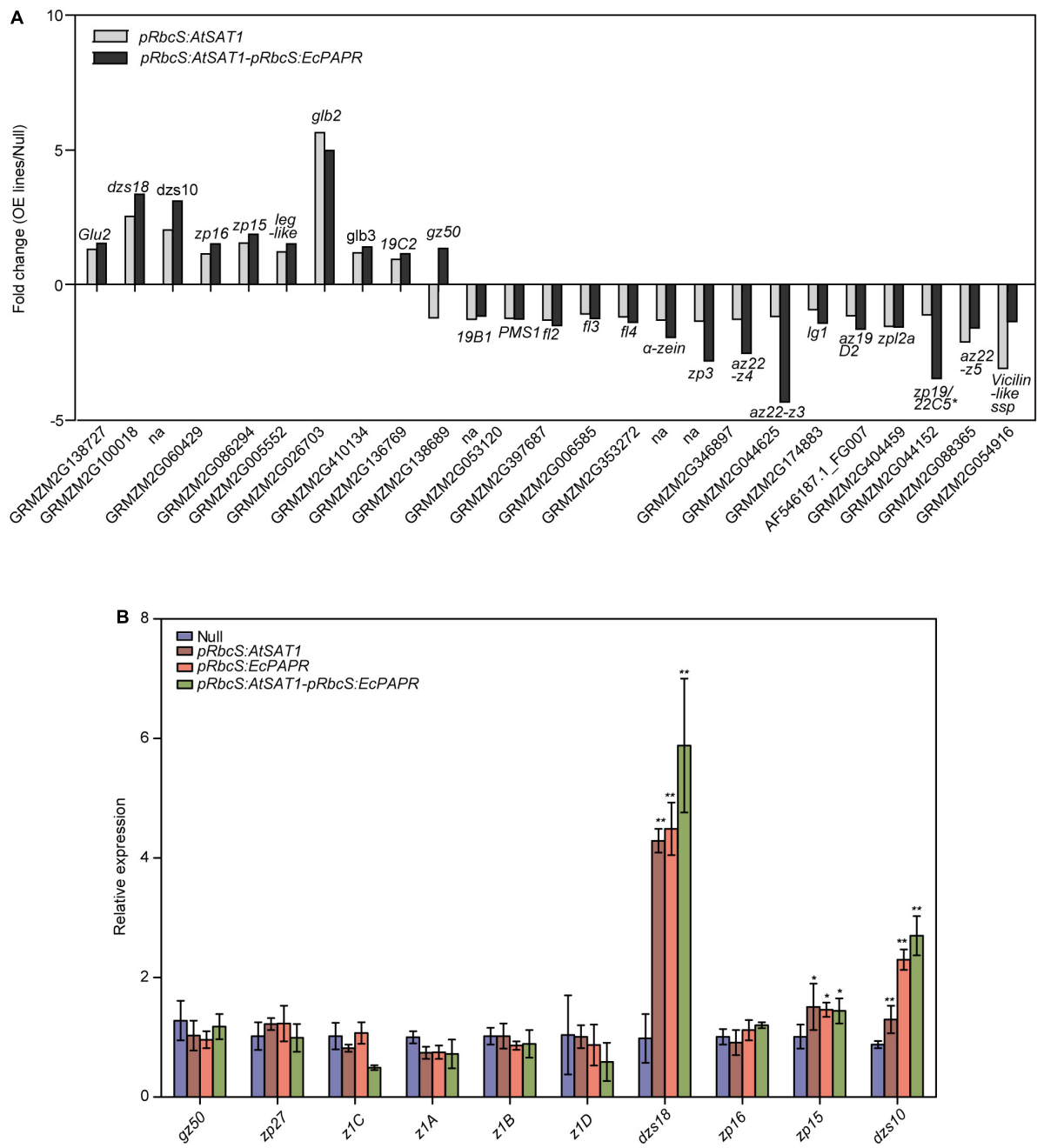


FIGURE 5 Comparison of gene expression patterns obtained using RNA-Seq and qRT-PCR. (A) Nutrient reservoir activity-related DEGs in *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR* compared with Null. Data were log₂-transformed and plotted in the bar graph. (B) Expression of *gz50*, *zpl27*, *z1C*, *z1A*, *z1B*, *z1D*, *dzs18*, *zpl6*, *zpl5*, and *dzs10* in the OE lines relative to Null. The data are from three biological replicates per sample and are presented as mean ± SD. Asterisks indicate a significant difference from Null. (Student's *t*-test, **p*-value < 0.05 and ***p*-value ≤ 0.001).

Upregulated transcription factors in endosperm mainly belong to EREB, bZIP, and HSFTF families (Table 3). Here, these TFs induced by increased MET flux into seeds might be involved in transcriptional or posttranscriptional regulation of *dzs18* and *dzs10* gene expression, thus affecting the 18- and 10-kDa δ-zein protein levels in the endosperm.

PPI analysis of upregulated DEGs in *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves and endosperm

A large PPI network that included most of the genes was constructed, containing 601 nodes. The MCODE clustering

algorithm was used to identify clusters in the PPI network. Using the MCODE plugin, four clusters (highly interconnected regions; **Figure 6**) in the networks were obtained with parameters set as follows: degree cut-off = 0.3, K-core = 4, and max depth = 100. A cluster is a complete n-node sub-graph, which means that within a sub-graph, each pair of nodes is connected by an edge (Zhuang et al., 2015). Detailed information for the gene symbols is listed in **Supplementary Table 6**. Most genes were upregulated in

the leaves of these clusters. The sulfur reduction-related genes (*APRL1*, *APRL2*, and *SiR1*) and zinc transporters (*ZT1* and *ZTG4*) were clustered in one cluster; *CCR3* (*serine/threonine-protein kinase CCR3*) is the central hub gene, which has been identified as an enzyme that dependent on serine residues for its activity (Reddy and Rajasekharan, 2007). *MDHAR* (*monodehydroascorbate reductase*), *Grx_l1* (*glutaredoxin subgroup III*), and *GST2* (*glutathione transferase 2*) were the hub genes of the other three clusters.

TABLE 2 List of nutrient reservoirs activity related DEGs.

id	Null vs <i>pRbcS:AtSAT1</i>		Null vs <i>pRbcS:AtSAT1-pRbcS:EcPAPR</i>		Description
	Fold change	P-value	Fold change	P-value	
Zm00001d020592	1.3	3.90E-01	1.5	6.84E-03	<i>Glutelin-2</i>
Zm00001d037436	2.5	2.42E-02	3.3	1.36E-06	<i>dzs18 - delta zein structural18</i>
Zm00001d045937	2.0	1.09E-01	3.1	1.87E-13	<i>dzs10</i>
Zm00001d005793	1.1	7.04E-01	1.5	8.53E-03	<i>zp16</i>
Zm00001d035760	1.5	2.49E-01	1.9	7.97E-05	<i>zp15 - zein protein, 15kDa</i>
Zm00001d011036	1.2	6.67E-01	1.5	3.67E-01	<i>legumin-like protein</i>
Zm00001d034413	5.6	3.23E-01	5.0	3.52E-01	<i>glb2 - globulin2</i>
Zm00001d038597	1.2	4.24E-01	1.4	2.35E-01	<i>glb3 - globulin3</i>
Zm00001d029782	–	–	1.1	4.61E-01	<i>19-kDa-zein (19C2)</i>
Zm00001d020591	–1.2	5.88E-01	1.3	2.31E-01	<i>gz50 - 50kD gamma zein</i>
Zm00001d048848	–1.3	1.34E-01	–1.1	3.45E-01	<i>19-kDa-zein (19B1)</i>
Zm00001d048850	–1.2	2.11E-01	–1.3	2.51E-01	<i>α-zein PMS1 precursor</i>
Zm00001d049243	–1.3	1.23E-01	–1.5	1.01E-01	<i>fl2 - floury2 (22kDa-zein 1</i>
Zm00001d009292	–1.1	7.92E-01	–1.2	5.66E-01	<i>fl3</i>
Zm00001d048851	–1.2	3.48E-01	–1.4	1.42E-01	<i>fl4 - floury4</i>
Zm00001d048816	–1.3	1.23E-01	–1.9	1.71E-05	<i>α-zein protein</i>
Zm00001d048817	–1.3	9.24E-02	–2.8	4.94E-08	<i>zp3 - zein protein3</i>
Zm00001d048812	–1.3	1.85E-01	–2.5	4.87E-04	<i>az22z4(az22z4 - 22kD alpha zein4)</i>
Zm00001d048809	–1.2	4.34E-01	–4.3	1.34E-06	<i>az22z3 - 22kD alpha zein3</i>
Zm00001d035700	–0.9	5.75E-01	–1.4	1.99E-02	<i>eg1 - legumin1</i>
Zm00001d030855	–1.1	7.28E-01	–1.6	1.65E-01	<i>az19D2 - alpha zein 19kDa D2</i>
Zm00001d048847	–1.5	1.13E-02	–1.6	2.82E-02	<i>zpl2a - zein polypeptidesL2a (zp12b)</i>
Zm00001d048813	–1.1	6.62E-01	–3.5	2.76E-05	<i>zp19/22C5*(zein protein SF4C candidate 5)</i>
Zm00001d048806	–2.1	3.22E-01	–1.9	4.79E-01	<i>az22z5 - 22kD alpha zein5</i>
Zm00001d029062	–3.1	1.14E-02	–1.3	5.40E-01	<i>Vicilin-like seed storage protein</i>
Zm00001d048818	–2.6	1.83E-02	–1.1	5.76E-01	unknown
Zm00001d049476	–1.5	4.64E-02	–1.7	3.98E-03	unknown
Zm00001d048176	–1.5	2.79E-01	–1.2	5.87E-01	unknown
Zm00001d048810	–1.6	1.51E-01	–3.3	9.86E-04	unknown
Zm00001d025059	–4.6	2.15E-03	–2.1	2.51E-01	unknown
Zm00001d048849	–1.3	2.10E-01	–1.1	4.63E-01	unknown
Zm00001d048852	–1.3	1.61E-01	–1.2	4.03E-01	unknown
Zm00001d019160	–1.7	3.17E-02	–1.5	5.67E-02	unknown
Zm00001d019162	–1.3	7.44E-01	–1.9	8.86E-02	unknown
Zm00001d048808	–3.5	4.19E-03	–1.4	6.44E-01	unknown
Zm00001d048219			–1.8	4.24E-01	<i>legumin-like protein</i>
Zm00001d004401			–4.6	1.04E-02	<i>germin-like protein subfamily 1 member 17</i>

* $P \leq 0.05$.

Identification of novel genes expressed in 18 DAP endosperm *via* PPI analysis of *pRbcS:AtSAT1-pRbcS:EcPAPR* up- and downregulated differentially expressed genes

To further explore the protein interaction in the 18 DAP endosperm of *pRbcS:AtSAT1-pRbcS:EcPAPR*, we applied a protein network-based approach to identify subnetworks that may provide new insights into the functions of pathways involved in high-MET storage protein rather than single genes. This reflects the aggregating behavior of genes connected in a PPI network (Goel et al., 2018). The network was binary, and all interactions were unweighted and undirected. A graph that included the majority of the genes containing 69 nodes was constructed (Figure 7) based on our analysis. The details of the gene symbols in the network are listed in Supplementary Table 8. The size of each node represents the degree index. The degree of its nodes indicates the number of interactions of a single node with all the other nodes. The top 10 hub genes included *HSP70* (70-kDa heat shock proteins), *PK* (Pyruvate kinase), *ARFB1B* (ADP-ribosylation factor B1B), *PPFα* (Pyrophosphate – fructose 6-phosphate 1-phosphotransferase alpha subunit), *CID11* (CTC-interacting domain 11, RNA-binding protein), *MADS52* (MADS transcription factor 52), *40SS28* (40S ribosomal protein S28), *TUB5* (beta tubulin5), *60SL31* (60S ribosomal protein L31), and *60SP2A* (60S acidic ribosomal protein P2A). The results revealed that *HSP70* was the centrality (hub) gene of the network, which functions actively in many aspects, such as protein folding, unfolding, regulation, targeting, aggregation and disaggregation, homeostasis, and degradation (Fernandez-Fernandez et al., 2017; Fernandez-Fernandez and Valpuesta, 2018; Rosenzweig et al., 2019).

Discussion

Overproduced sulfur-related compounds in *pRbcS:AtSAT1-pRbcS:EcPAPR* leaves had negative effects on plant growth

Increased Met and Cys in *pRbcS:AtSAT1-pRbcS:EcPAPR* indicated a higher sulfur content in the plants than wild type. Increased S reduction has profound consequences for the synthesis of all S-metabolites, including sulfur amino acids Met, Cys, and especially GSH and changed cell processes, such as oxidation–reduction, amino acid transport, and TF expression levels. The *pRbcS:AtSAT1-pRbcS:EcPAPR* plants grew slowly, were weaker, and had less biomass, smaller ears, and fewer kernels per ear, which is similar to *PaAPR* and *EcPAPR* co-overexpression lines (Martin et al., 2005).

TABLE 3 TFs up regulated in 18 DAP endosperm of *pRbcS:AtSAT1* and *pRbcS:AtSAT1-pRbcS:EcPAPR*.

Up-down regulation	Gene ID (TF name or family)	
	<i>pRbcS:AtSAT1</i>	<i>pRbcS:AtSAT1-pRbcS:EcPAPR</i>
Up-regulated TFs in 18 DAP endosperm	Zm00001d019216 (<i>ereb64</i>)	Zm00001d019216 (<i>ereb64</i>)
	Zm00001d022461 (<i>ereb200</i>)	Zm00001d000179 (<i>ereb1</i>) Zm00001d024324 (<i>ereb54</i>) Zm00001d023332 (<i>wrky63</i>) Zm00001d043782 (<i>ereb126</i>) Zm00001d002762 (<i>ereb198</i>) Zm00001d026271 (<i>ereb205</i>)
	Zm00001d002143 (<i>bzip27</i>)	Zm00001d002143 (<i>bzip27</i>)
	Zm00001d009160 (<i>Trihelix</i>)	Zm00001d009160 (<i>Trihelix</i>)
	Zm00001d029506 (<i>lbd5</i>)	Zm00001d029506 (<i>lbd5</i>)
	Zm00001d038717 (<i>lbd33</i>)	
	Zm00001d010751 (<i>lbd38</i>)	
	Zm00001d052738 (<i>hsf7</i>)	Zm00001d052738 (<i>hsf7</i>)
	Zm00001d010812 (<i>hsf16</i>)	Zm00001d026094 (<i>hsf20</i>)
	Zm00001d023669 (<i>naftf67</i>)	Zm00001d023332 (<i>heat shock</i>) Zm00001d043921 (<i>nactf82</i>) Zm00001d020492 (<i>wrky53</i>) Zm00001d022099 (<i>ca3p4</i>) Zm00001d028930 (<i>myb75</i>) Zm00001d017147 (<i>dbb6</i>) Zm00001d026536 (<i>C3H</i>) Zm00001d024200 (<i>CO-like</i>)
		Zm00001d041958 (<i>WRKY</i>)
Down-regulated TFs in 18 DAP endosperm	Zm00001d012757 (<i>GATA</i>)	Zm00001d012757 (<i>GATA</i>)
	Zm00001d026351 (<i>HD-ZIP</i>)	Zm00001d040651 (<i>ERF</i>)
	Zm00001d033215 (<i>NF-YA</i>)	Zm00001d039913 (<i>MADS54</i>)
	Zm00001d026447 (<i>ERF</i>)	Zm00001d031620 (<i>MIKC</i>)
	Zm00001d047967 (<i>bZip</i>)	Zm00001d040301 (<i>C3H</i>) Zm00001d021019 (<i>bHLH</i>) Zm00001d037098 (<i>bHLH</i>) Zm00001d017614 (<i>MIKC</i>) Zm00001d042560 (<i>LBD</i>) Zm00001d046755 (<i>ARR-B</i>) Zm00001d040362 (<i>Dof</i>) Zm00001d002718 (<i>GRAS</i>) Zm00001d015381 (<i>M-type</i>) Zm00001d041489 (<i>HD-ZIP</i>) Zm00001d030995 (<i>bZIP</i>) Zm00001d040536 (<i>bHLH</i>)

From our transcriptome analysis, this might occur because S overproduction causes leaf cells to overaccumulate reaction sulfur species (RSS) in plants. RSS in the form of persulfidated cysteines (Cys-S-S) is produced endogenously and co-translationally introduced into proteins (Olson, 2020). *pRbcS:AtSAT1-pRbcS:EcPAPR* had plant cellular damage and cell component changes due to high levels of S-compounds, such as RSS. Oxidative stress induced by RSS can lead to cell

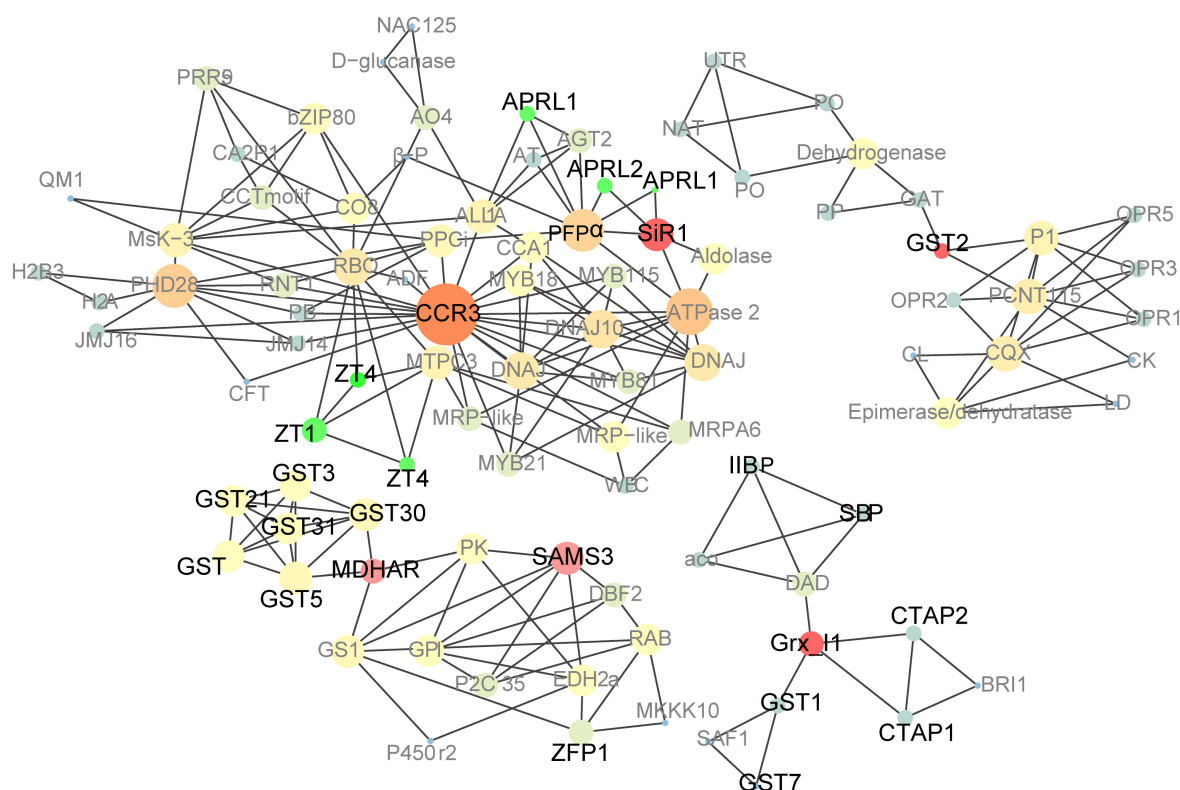


FIGURE 6
Co-expression network analysis of up-regulated genes in *pRbcS:AtSAT1-pRbcS:EcPAPR* mature leaves and 18 DAP endosperm. Cycle nodes represent genes, and the sizes of the nodes represent the power of the interrelation among the nodes. Edges between two nodes represent interactions between genes. The more edges of a gene, the more genes are connected to it, and the more central role it has within the network.

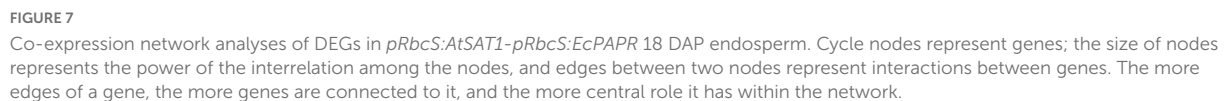
death and tissue injury (Foyer and Noctor, 2011; Olson, 2021). Plants have evolved antioxidant mechanisms to deal with RSS, in which many genes, such as *S-transferase genes* (GSTs), ascorbate peroxidase genes, and glutaredoxins (GRXs), are involved (Xie et al., 2018). In this study, several GST genes were upregulated in *pRbcS:AtSAT1-pRbcS:EcPAPR*, such as *GST2*, *GST5*, *GST6*, *GST7*, *GST21*, *GST22*, *GST30*, *GST34*, and *GST37*, which mainly catalyze the conjugation of GSH onto xenobiotics, and some have GSH-dependent peroxidase activity against H₂O₂, H₂S, and organic peroxides (Dixon et al., 2009; Tossounian et al., 2018; Olson, 2020). Several GST-encoding genes are strongly induced by oxidative stress (Wachter et al., 2005), thus helping plants cope with biotic and abiotic stress. This indicates that RSS induced by sulfur overproduction may be the main reason for stunting and dwarf phenotypes in *pRbcS:AtSAT1-pRbcS:EcPAPR* plants.

RSS signals *via* oxidation reactions with protein cysteine sulfur, and they produce identical effector responses, such as GSH. GSH can modulate reactive oxygen species by oxidizing the cysteine residues of transcription factors and signaling molecules (Lu, 2009). Many genes related to GSH synthase were upregulated in *pRbcS:AtSAT1-pRbcS:EcPAPR*, including *Glutamate synthase 1* (GS1), which is the key

enzyme in N assimilation involved in glutamine synthesis (Seger et al., 2015); *glutamate-cysteine ligase B* (GCL), which catalyzes GSH biosynthesis (Musgrave et al., 2013); and *gamma-glutamylcysteine synthetase1* (GSH1), which is conjugated with CYS to form γ -glutamylcysteine (Lu, 2013). GSH1 overexpression has a negative effect on tobacco growth (Creissen et al., 1999). The negative effects of GSH accumulation also explain this phenotype (Martin et al., 2005).

Met production in maize leaves and transport to seeds

Several genes related to Met metabolism were upregulated in *pRbcS:AtSAT1-pRbcS:EcPAPR*, including *Homocysteine S-methyltransferase 1* (HMT1), which catalyzes SMM to be reconverted to Met (Mudd and Datko, 1990); *S-adenosylmethionine synthetase* (SAMS3), which is formed by MET adenosylation and is the precursor plant of certain polyamines and the plant hormone ethylene (Roje, 2006); and *O-methyltransferase* (OMT), which is dependent on S-adenosyl-l-methionine and can catalyze a variety of secondary metabolites (Liu et al., 2018). Some amino acid transporter



The transport of amino acids into plant cells is known to involve a range of proton-coupled symporters with various substrate specificities, some of which overlap (Rentsch et al., 1998). To date, SMM appears not to have been tested as a substrate. *A priori*, it could be a substrate for the known general amino acid or basic amino acid transporter (Chen and Bush, 1997; Rentsch et al., 1998); alternatively, there could be an SMM-specific transporter similar to the SAM permease

Therefore, engineering the SMM-related transporter would also increase the MET flux into the seeds. The combination of S reduction key enzymes and SMM transporters is another approach for super-high MET maize breeding.

GSH, together with phytochelatin, is known to contribute to heavy metal detoxification and a range of stress responses (Cobbett, 2000; Lu, 2013). Many heavy metal-related genes were

Author contributions

CL and XX was conceived and designed by the experiment. BH, ZP, and LW collected the phenotypes and performed the data analysis. XX and CL was written by the manuscript. LT was edited the manuscript. All authors have read and approved the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (31901564, 32170639, and 21242010).

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.969763/full#supplementary-material>

References

- Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169. doi: 10.1093/bioinformatics/btu638
- Anders, S., and Huber, W. (2012). *Differential expression of RNA-Seq data at the gene level—the DESeq package*. Heidelberg: European Molecular Biology Laboratory (EMBL), 10.
- Bader, G. D., and Hogue, C. W. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4:2. doi: 10.1186/1471-2105-4-2
- Bagga, S., Potenza, C., Ross, J., Martin, M. N., Leustek, T., and Sengupta-Gopalan, C. (2005). A transgene for high methionine protein is posttranscriptionally regulated by methionine. *In Vitro Cell. Dev. Biol. Plant* 41, 731–741. doi: 10.1079/IVP2005709
- Benner, M. S., Phillips, R. L., Kiriara, J. A., and Messing, J. W. (1989). Genetic analysis of methionine-rich storage protein accumulation in maize. *Theor. Appl. Genet.* 78, 761–767. doi: 10.1007/BF00266655
- Bourgis, F., Roje, S., Nuccio, M. L., Fisher, D. B., Tarczynski, M. C., Li, C., et al. (1999). S-methylmethionine plays a major role in phloem sulfur transport and is synthesized by a novel type of methyltransferase. *Plant Cell* 11, 1485–1498. doi: 10.1105/tpc.11.8.1485
- Brunold, C., and Rennenberg, H. (1997). Regulation of sulfur metabolism in plants: First molecular approaches. *Prog. Bot.* 58, 164–186.
- Chaudhuri, S., and Messing, J. (1994). Allele-specific parental imprinting of *dzrl*, a posttranscriptional regulator of zein accumulation. *Proc. Natl. Acad. Sci. U.S.A.* 91, 4867–4871. doi: 10.1073/pnas.91.11.4867
- Chen, L., and Bush, D. R. (1997). LHT1, a lysine- and histidine-specific amino acid transporter in Arabidopsis. *Plant Physiol.* 115, 1127–1134.
- Cobbett, C. S. (2000). Phytochelatin and their roles in heavy metal detoxification. *Plant Physiol.* 123, 825–832.
- Creissen, G., Firmin, J., Fryer, M., Kular, B., Leyland, N., Reynolds, H., et al. (1999). Elevated glutathione biosynthetic capacity in the chloroplasts of transgenic tobacco plants paradoxically causes increased oxidative stress. *Plant Cell* 11, 1277–1292. doi: 10.1105/tpc.11.7.1277
- Dixon, D. P., Hawkins, T., Hussey, P. J., and Edwards, R. (2009). Enzyme activities and subcellular localization of members of the Arabidopsis glutathione transferase superfamily. *J. Exp. Bot.* 60, 1207–1218. doi: 10.1093/jxb/ern365
- Fernandez-Fernandez, M. R., Gragera, M., Ochoa-Ibarrola, L., Quintana-Gallardo, L., and Valpuesta, J. M. (2017). Hsp70 - a master regulator in protein degradation. *FEBS Lett.* 591, 2648–2660. doi: 10.1002/1873-3468.12751
- Fernandez-Fernandez, M. R., and Valpuesta, J. M. (2018). Hsp70 chaperone: A master player in protein homeostasis. *F1000Res* 7:F1000FacultyRev-1497. doi: 10.12688/f1000research.15528.1
- Fisher, J. M. A. H. (1991). Maternal effect on high methionine levels in hybrid corn. *J. Biotechnol.* 21, 229–238.
- Foyer, C. H., and Noctor, G. (2011). Ascorbate and glutathione: The heart of the redox hub. *Plant Physiol.* 155, 2–18. doi: 10.1104/pp.110.167569
- Goel, P., Sharma, N. K., Bhuria, M., Sharma, V., Chauhan, R., Pathania, S., et al. (2018). Transcriptome and Co-Expression Network Analyses Identify Key Genes Regulating Nitrogen Use Efficiency in Brassica juncea L. *Sci. Rep.* 8:7451. doi: 10.1038/s41598-018-25826-6
- Heo, J. O., Chang, K. S., Kim, I. A., Lee, M. H., Lee, S. A., Song, S. K., et al. (2011). Funneling of gibberellin signaling by the GRAS transcription regulator scarecrow-like 3 in the Arabidopsis root. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2166–2171. doi: 10.1073/pnas.1012215108
- International, A. (1997). *Official Methods of Analysis of AOAC International*, 16th Edn. Gaithersburg, MD: AOAC International.
- Kawachi, M., Kobae, Y., Mori, H., Tomioka, R., Lee, Y., and Maeshima, M. (2009). A mutant strain Arabidopsis thaliana that lacks vacuolar membrane zinc transporter MTP1 revealed the latent tolerance to excessive zinc. *Plant Cell Physiol.* 50, 1156–1170. doi: 10.1093/pcp/pcp067
- Kim, Y. Y., Choi, H., Segami, S., Cho, H. T., Martinoia, E., Maeshima, M., et al. (2009). AtHMA1 contributes to the detoxification of excess Zn(II) in Arabidopsis. *Plant J.* 58, 737–753. doi: 10.1111/j.1365-3113X.2009.03818.x
- Lai, J., and Messing, J. (2002). Increasing maize seed methionine by mRNA stability. *Plant J.* 30, 395–402. doi: 10.1046/j.1365-3113x.2001.01285.x
- Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9:559. doi: 10.1186/1471-2105-9-559
- Lending, C. R., and Larkins, B. A. (1989). Changes in the zein composition of protein bodies during maize endosperm development. *Plant Cell* 1, 1011–1023. doi: 10.1105/tpc.1.10.1011
- Li, C., Yue, Y., Chen, H., Qi, W., and Song, R. (2018). The ZmZIP22 Transcription Factor Regulates 27-kD gamma-Zein Gene Transcription during Maize Endosperm Development. *Plant Cell* 30, 2402–2424. doi: 10.1105/tpc.18.00422

- Liu, Z., Fan, M., Li, C., and Xu, J. H. (2018). Dynamic gene amplification and function diversification of grass-specific O-methyltransferase gene family. *Genomics* 111, 687–695. doi: 10.1016/j.ygeno.2018.04.005
- Lu, S. C. (2009). Regulation of glutathione synthesis. *Mol. Aspects Med.* 30, 42–59. doi: 10.1016/j.mam.2008.05.005
- Lu, S. C. (2013). Glutathione synthesis. *Biochim. Biophys. Acta.* 1830, 3143–3153. doi: 10.1016/j.bbagen.2012.09.008
- Mach, J. (2018). Corn ChIPs and RNA-seq: Researchers Dip into Advanced Tools and Resources to Examine bZIP Transcription Factor Function in the Maize Endosperm. *Plant Cell* 30, 2641–2642. doi: 10.1105/tpc.18.00882
- Martin, M. N., Tarczynski, M. C., Shen, B., and Leustek, T. (2005). The role of 5'-adenylylsulfate reductase in controlling sulfate reduction in plants. *Photosynth. Res.* 86, 309–323. doi: 10.1007/s11120-005-9006-z
- Mizrachi, E., Hefer, C. A., Ranik, M., Joubert, F., and Myburg, A. A. (2010). *De novo* assembled expressed gene catalog of a fast-growing *Eucalyptus* tree produced by Illumina mRNA-Seq. *BMC Genomics* 11:681. doi: 10.1186/1471-2164-11-681
- Moreno, I., Norambuena, L., Maturana, D., Toro, M., Vergara, C., Orellana, A., et al. (2008). AtHMA1 is a thapsigargin-sensitive Ca²⁺/heavy metal pump. *J. Biol. Chem.* 283, 9633–9641. doi: 10.1074/jbc.M800736200
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Mudd, S. H., and Datko, A. H. (1990). The S-Methylmethionine Cycle in Lemna paucicostata. *Plant Physiol.* 93, 623–630. doi: 10.1104/pp.93.2.623
- Musgrave, W. B., Yi, H., Kline, D., Cameron, J. C., Wignes, J., Dey, S., et al. (2013). Probing the origins of glutathione biosynthesis through biochemical analysis of glutamate-cysteine ligase and glutathione synthetase from a model photosynthetic prokaryote. *Biochem. J.* 450, 63–72. doi: 10.1042/BJ20121332
- Olson, K. R. (2020). Reactive oxygen species or reactive sulfur species: Why we should consider the latter. *J. Exp. Biol.* 223:jeb196352. doi: 10.1242/jeb.196352
- Olson, K. R. (2021). The biological legacy of sulfur: A roadmap to the future. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 252:110824. doi: 10.1016/j.cbpa.2020.110824
- Planta, J., Xiang, X., Leustek, T., and Messing, J. (2017). Engineering sulfur storage in maize seed proteins without apparent yield loss. *Proc. Natl. Acad. Sci. U.S.A.* 114, 11386–11391. doi: 10.1073/pnas.1714805114
- Ralf Weigel, T. S., Krawczyk, S., and Gat, C. (2002). “bZIP factors TGA2 and TGA6 interact with SCRARECROW-LIKE protein 14 (SCL14),” in *13th International Conference On Arabidopsis Research*, (Seville).
- Reddy, M. M., and Rajasekharan, R. (2007). Serine/threonine/tyrosine protein kinase from *Arabidopsis thaliana* is dependent on serine residues for its activity. *Arch. Biochem. Biophys.* 460, 122–128. doi: 10.1016/j.abb.2007.01.003
- Rentsch, D., Boorer, K. J., and Frommer, W. B. (1998). Structure and function of plasma membrane amino acid, oligopeptide and sucrose transporters from higher plants. *J. Membr. Biol.* 162, 177–190. doi: 10.1007/s002329900355
- Roje, S. (2006). S-Adenosyl-L-methionine: Beyond the universal methyl group donor. *Phytochemistry* 67, 1686–1698. doi: 10.1016/j.phytochem.2006.04.019
- Rosenzweig, R., Nilligoda, N. B., Mayer, M. P., and Bukau, B. (2019). The Hsp70 chaperone network. *Nat. Rev. Mol. Cell. Biol.* 20, 665–680. doi: 10.1038/s41580-019-0133-3
- Rouillon, A., Surdin-Kerjan, Y., and Thomas, D. (1999). Transport of sulfonium compounds. Characterization of the s-adenosylmethionine and s-methylmethionine permeases from the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 274, 28096–28105. doi: 10.1074/jbc.274.40.28096
- Seger, M., Gebril, S., Tabilona, J., Peel, A., and Sengupta-Gopalan, C. (2015). Impact of concurrent overexpression of cytosolic glutamine synthetase (GS1) and sucrose phosphate synthase (SPS) on growth and development in transgenic tobacco. *Planta* 241, 69–81. doi: 10.1007/s00425-014-2165-4
- Seigneurin-Berny, D., Gravot, A., Auroy, P., Mazard, C., Kraut, A., Finazzi, G., et al. (2006). HMA1, a new Cu-ATPase of the chloroplast envelope, is essential for growth under adverse light conditions. *J. Biol. Chem.* 281, 2882–2892. doi: 10.1074/jbc.M508333200
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Tabe, 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
- Takahashi, H., Kopriva, S., Giordano, M., Saito, K., and Hell, R. (2011). Sulfur assimilation in photosynthetic organisms: Molecular functions and regulations of transporters and assimilatory enzymes. *Annu. Rev. Plant Biol.* 62, 157–184. doi: 10.1146/annurev-arplant-042110-103921
- Team, R. C. (2014). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.
- Thompson, G. A., and Larkins, B. A. (1994). “Characterization of zein genes and their regulation in maize endosperm,” in *The Maize Handbook*, eds M. Freeling and V. Walbot (New York, NY: Springer-Verlag), 639–647.
- Tossounian, M. A., Wahni, K., Van Molle, L., Vertommen, D., Astolfi Rosado, L., and Messens, J. (2018). Redox regulated methionine oxidation of *Arabidopsis thaliana* glutathione transferase Phi9 induces H-site flexibility. *Protein Sci.* 28, 56–67. doi: 10.1002/pro.3440
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578. doi: 10.1038/nprot.2012.016
- Wachter, A., Wolf, S., Steininger, H., Bogs, J., and Rausch, T. (2005). Differential targeting of GSH1 and GSH2 is achieved by multiple transcription initiation: Implications for the compartmentation of glutathione biosynthesis in the Brassicaceae. *Plant J.* 41, 15–30. doi: 10.1111/j.1365-3113X.2004.02269.x
- Williams, L. E., and Mills, R. F. (2005). P(1B)-ATPases—an ancient family of transition metal pumps with diverse functions in plants. *Trends Plant Sci.* 10, 491–502. doi: 10.1016/j.tplants.2005.08.008
- Wu, Y., Goettel, W., and Messing, J. (2009). Non-Mendelian regulation and allelic variation of methionine-rich delta-zein genes in maize. *Theor. Appl. Genet.* 119, 721–731. doi: 10.1007/s00122-009-1083-5
- Wu, Y., and Messing, J. (2010). RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. *Plant Physiol.* 153, 337–347. doi: 10.1104/pp.110.154690
- Wu, Y., Wang, W., and Messing, J. (2012). Balancing of sulfur storage in maize seed. *BMC Plant Biol.* 12:77. doi: 10.1186/1471-2229-12-77
- Xiang, X., Wu, Y., Planta, J., Messing, J., and Leustek, T. (2017). Overexpression of serine acetyltransferase in maize leaves increases seed-specific methionine-rich zeins. *Plant Biotechnol. J.* 6, 1057–1067.
- Xiang, X., Wu, Y., Planta, J., Messing, J., and Leustek, T. (2018). Overexpression of serine acetyltransferase in maize leaves increases seed-specific methionine-rich zeins. *Plant Biotechnol. J.* 16, 1057–1067. doi: 10.1111/pbi.12851
- Xie, R., Pan, X., Zhang, J., Ma, Y., He, S., Zheng, Y., et al. (2018). Effect of salt-stress on gene expression in citrus roots revealed by RNA-seq. *Funct. Integr. Genomics* 18, 155–173. doi: 10.1007/s10142-017-0582-8
- Zhuang, D. Y., Jiang, L., He, Q. Q., Zhou, P., and Yue, T. (2015). Identification of hub subnetwork based on topological features of genes in breast cancer. *Int. J. Mol. Med.* 35, 664–674. doi: 10.3892/ijmm.2014.2057



OPEN ACCESS

EDITED BY
Jihong Hu,
Northwest A&F University, China

REVIEWED BY
Yin Li,
Sun Yat-sen University, China
Jianxin Shi,
Shanghai Jiao Tong University, China

*CORRESPONDENCE
Chuanzhi Zhao
chuanzhiz@126.com
Yindong Zhang
23300558@163.com

†These authors have contributed
equally to this work

SPECIALTY SECTION
This article was submitted to
Plant Bioinformatics,
a section of the journal
Frontiers in Plant Science

RECEIVED 12 July 2022
ACCEPTED 29 August 2022
PUBLISHED 16 September 2022

CITATION
Zhang K, Ma J, Gangurde SS, Hou L,
Xia H, Li N, Pan J, Tian R, Huang H,
Wang X, Zhang Y and Zhao C (2022)
Targeted metabolome analysis reveals
accumulation of metabolites in testa
of four peanut germplasms.
Front. Plant Sci. 13:992124.
doi: 10.3389/fpls.2022.992124

COPYRIGHT
© 2022 Zhang, Ma, Gangurde, Hou,
Xia, Li, Pan, Tian, Huang, Wang, Zhang
and Zhao. 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(s)
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.

Targeted metabolome analysis reveals accumulation of metabolites in testa of four peanut germplasms

Kun Zhang^{1,2,3†}, Jing Ma^{2,4†}, Sunil S. Gangurde^{5,6}, Lei Hou^{2,4},
Han Xia^{2,4}, Nana Li², Jiaowen Pan², Ruizheng Tian²,
Huailing Huang^{2,4}, Xingjun Wang^{2,4}, Yindong Zhang^{1,7*} and
Chuanzhi Zhao^{2,4*}

¹College of Tropical Crops, Hainan University, Haikou, China, ²Institute of Crop Germplasm Resources (Institute of Biotechnology), Shandong Academy of Agricultural Sciences, Shandong Provincial Key Laboratory of Crop Genetic Improvement, Ecology and Physiology, Jinan, China, ³College of Agricultural Science and Technology, Shandong Agriculture and Engineering University, Jinan, China, ⁴College of Life Sciences, Shandong Normal University, Jinan, China, ⁵Crop Protection and Management Research Unit, USDA-ARS, Tifton, GA, United States, ⁶Department of Plant Pathology, University of Georgia, Tifton, GA, United States, ⁷Hainan Academy of Agricultural Sciences, Haikou, China

Cultivated peanut (*Arachis hypogaea* L.) is an important source of edible oil and protein. Peanut testa (seed coat) provides protection for seeds and serves as a carrier for diversity metabolites necessary for human health. There is significant diversity available for testa color in peanut germplasms. However, the kinds and type of metabolites in peanut testa has not been comprehensively investigated. In this study, we performed metabolite profiling using UPLC-MS/MS for four peanut germplasm lines with different testa colors, including pink, purple, red, and white. A total of 85 metabolites were identified in four peanuts. Comparative metabolomics analysis identified 78 differentially accumulated metabolites (DAMs). Some metabolites showed significant correlation with other metabolites. For instance, proanthocyanidins were positively correlated with cyanidin 3-O-rutinoside and malvin, and negatively correlated with pelargonidin-3-glucoside. We observed that the total proanthocyanidins are most abundant in pink peanut variety WH10. The red testa accumulated more isoflavones, flavonols and anthocyanidins compared with that in pink testa. These results provided valuable information about differential accumulation of metabolites in testa with different color, which are helpful for further investigation of the molecular mechanism underlying biosynthesis and accumulation of these metabolites in peanut.

KEYWORDS

testa color, flavonoids, LC-MS/MS, metabolome profiling, metabolic pathway

Introduction

Cultivated peanut (*Arachis hypogaea* L.) is rich in protein, oil, and nutrients. Peanut is an important oilseed crop of Asia, Africa, and Americas (FAOSTAT 2020¹). The edible portion of the peanut consists of both kernel and the seed coat also called as peanut testa (Christman et al., 2018). Previous studies reported that the peanut testa is rich in nutrients, including 12.3% protein, 16.6% oil, 2.8% ash, and approximately 140~150 mg/g phenolic compounds (Nepote et al., 2007). Phenolic compounds are usually enriched on the exosphere of plant tissue to protect the internal cells. Peanut testa is rich in such phenolic compounds (Ma et al., 2014). In recent years, with the deepening of research on the important role of plant polyphenols in human health, more attentions have been paid to the nutritional value and function of peanut testa (Yu et al., 2006; Constanza et al., 2012; Liu et al., 2020; Oliveira et al., 2021).

Flavonoids and non-flavonoids are the major polyphenols in peanut testa. Flavonoids have the basic structure which consists of a C6-C3-C6 carbon skeleton comprising two 6-carbon benzene rings (rings A and B) linked by a 3-carbon heterocyclic ring (ring C) (Nabavi et al., 2020). Based on the degree of oxidation of the ring C and the number of methyl or hydroxyl groups on the rings A and B, flavonoids are mainly divided into 12 subgroups: flavanones, flavones, isoflavones, flavonols, anthocyanidins, proanthocyanidins, chalcones, dihydrochalcones, dihydroflavonols, coumarins, aurones, and phlobaphenes (Winkel-Shirley, 2001; Yu et al., 2005; Sasaki and Nakayama, 2015). With the diversity of molecular polymerization and modification (glycosylation, acylation, and others), there are more than 9,000 kinds of flavonoids present in higher plants (Tanaka et al., 2009; Noda et al., 2017; Sun et al., 2020). Non-flavonoids have the basic structural skeletons different from that of flavonoids, such as stilbenes (C6-C2-C6), phenolic acids (C6-C1), and hydroxycinnamic acids (C6-C3), all of which have dietary and medical significance. Different from the flavonoid pathway, the non-flavonoids are formed through the shikimic and phenylpropanoid pathways (Crozier et al., 2009). The scope of flavonoids and non-flavonoids is still controversial. For example, one view is that stilbene belongs to flavonoids. And resveratrol, an important member of stilbenes, is considered to have great potential in the treatment of cancer (Winkel-Shirley, 2001; Siroerol et al., 2016).

As one of the most important secondary metabolites in plants, flavonoids play an important role in plant growth, development and resistance to biotic, and abiotic stresses. Anthocyanidins are water-soluble pigments and important member of flavonoids in plants. They are important coloring substances of plant tissues, such as fruits and flowers with red, orange, blue, and purple colors (Grotewold, 2006;

Dong et al., 2019), which can attract animals and insects, thus promoting pollination (Bradshaw and Schemske, 2003). Moreover, anthocyanidins have reactive oxygen species (ROS) scavenging ability and protect plants against damage from biotic and abiotic stresses (Cavauiuolo et al., 2013). Proanthocyanidins are the flavonoids with stronger antioxidant capacity than anthocyanidins, and their content is as high as 17% (w/w) of the dry weight of peanut testa (Karchesy and Hemingway, 1986; Park et al., 2011). Many other flavonoids, as phytoalexins or antioxidants, have properties similar to anthocyanidins and play important roles in resistance to drought stress, cold, ultraviolet radiations, and resistance to microorganisms, pathogens and insects (Iwashina, 2003; Pourcel et al., 2007; Zhang et al., 2020). Flavonoids can provide protection against oxidative stress, which has been implicated in some human diseases (Karadag et al., 2009). For example, flavonoid play a positive role in the treatment of atherosclerosis, diabetes mellitus, chronic inflammation, and some types of cancers in humans (Pinent et al., 2004; Osakabe and Yamagishi, 2009; Liu et al., 2020; Oliveira et al., 2021). Due to these bioactivities for flavonoids in peanut testa, attempts have been made to use them in some form as functional food ingredients or purified drug (Francisco and Resurreccion, 2009; Hathorn and Sanders, 2012; Christman et al., 2018). Flavonoids are generated from phenylalanine through the phenylpropanoid pathway, which has been clearly elucidated in model plants. With a series of enzyme reactions, metabolites flow to flavonols, flavanols, isoflavones, anthocyanidins and other branches, forming flavonoids substance with great differences in structure and content (Dong and Lin, 2021).

Testa color is an important qualitative trait of peanut. There are abundant phenotypic variations available in peanut testa color including pink, red, purple (black), white, tan, and multicolor, pink is most common in peanut germplasm. The composition and content of anthocyanidins are closely related to the color intensity of peanut testa. For instance, cyanidin-3-O-sambubioside which is a main anthocyanidin presents in black testa peanut cultivars (Kuang et al., 2017). With the completion of genome sequencing of cultivated peanut, great progress has been made in the research on the molecular mechanism of peanut seed coat color formation and gene identification. The candidate gene *AhTc1* that controls the color of peanut purple testa has been mapped on chromosome 10 and encodes for R2R3-MYB transcription factor (Zhao et al., 2019). Two independently inherited genes controlling peanut red testa, *AhRt1* and *AhRt2*, were identified on chromosomes 3 and 12, respectively (Chen et al., 2021; Zhang et al., 2022). Some other candidate genes and critical pathway were also predicted by combining transcriptome and metabolome methods between testa of two different colors, which enriched the study of gene function and the relationship between genes and metabolites in peanut testa (Wan et al., 2020; Xue et al., 2021). However, there are very few reports on the metabolomics

¹ <http://www.fao.org/faostat/en/#data/>

studies of multiple testa colors of peanut. The metabolic pathways involved in flavonoid biosynthesis and molecular mechanism underlying testa color formation in peanut is still unclear.

Therefore, in this study, flavonoids content in various peanut testa color differences were investigated with the help of metabolomics profiling of four different colored peanut testas. The results provided insights on characteristics and direction of flavonoids accumulation in different colored peanut testa and supported for improving cultivated peanut with target color and flavonoid quality.

Materials and methods

Plant materials and treatments

The four peanut cultivars namely Yuhua 29 (YH29) with purple testa, Zhonghua 12 (ZH12) with red testa, Weihua 10 (WH10) with pink testa, Kainongbai (KNB) with white testa were planted at Jiyang Experimental Station of Shandong Academy of Agricultural Sciences (SAAS), Shandong, China (36°58'34.53" N, 116°59'1.29" E) during 2020. The peanut testa samples were peeled carefully and collected when the seeds were on the period of 70 days after pegging (DAP 70). Three biological replicates were performed for each peanut cultivar and each replicate contained the testa of 10 seeds from many plants of this variety. Then samples were rapidly frozen in liquid nitrogen and stored at -80°C .

Chemicals and reagents

HPLC-grade methanol, water, acetonitrile was purchased from Thermo Scientific (Rockford, IL, USA), AR-grade chloroform was purchased from Titan (Shanghai, China), HPLC-grade flavonoid standards was purchased from Sigma-Aldrich (Shanghai) Trading Co., Ltd., (Shanghai, China) and Yuanye (Shanghai, China). All the other chemicals were of analytical grade.

Sample extraction

The powdered sample (50 mg) was extracted with $600^{\circ}\mu\text{L}$ of aqueous methanol (v:v = 1:2, containing succinic acid-2,2,3,3-d₄, 50°ng/mL) using a rapid grinder (JXFSTPRP-24/32; Shanghai, China) at a frequency of 60 Hz for 2 min, followed by 20 min of sonication at 4°C (SB-5200DT, Ningbo, China). Subsequently, the samples were centrifuged at 13,000 rpm for 10 min at 4°C . $500^{\circ}\mu\text{L}$ supernatant was transferred into new centrifuge tube. The residue added $400^{\circ}\mu\text{L}$ methanol, extracted and centrifuged through the same steps as above, got $300^{\circ}\mu\text{L}$

supernatant. The two extraction supernatants were mixed to obtain a total of $800^{\circ}\mu\text{L}$. Took $200^{\circ}\mu\text{L}$ of supernatant and evaporated, then redissolved with $200^{\circ}\mu\text{L}$ aqueous methanol (v:v = 18:7, containing standard l-2-chlorophenylalanine, 10°ng/mL). Centrifuged at 13,000 rpm for 5 min at 4°C , the supernatants were filtered using $0.22\text{-}\mu\text{m}$ organic phase pinhole filter. Transfer them to brown LC injection vial and store at -80°C until LC-MS/MS analysis.

Flavonoids-metabolites detection and multiple reaction monitoring

Metabolite quantification was analyzed by multiple reaction detection (MRM) mode of triple quadrupole mass spectrometry (MS). Flavonoids-metabolites identification and quantification was carried out by Shanghai Luming Biological Technology Co., Ltd., (Shanghai, China) using UPLC-ESI-MS/MS system (UPLC, AB ExionLC, Applied Biosystems Sciex; MS, Qtrap 6500+, Applied Biosystems Sciex). The analytical conditions were as follows: HPLC: column, Waters UPLC HSS T3 ($100^{\circ}\times 2.1\text{ mm}$, $1.8\text{ }\mu\text{m}$); solvent system, mobile phase A (0.01% formic acid); mobile phase B (acetonitrile); gradient program, 100:0 V/V at 0 min, 5:95 V/V at 11.0 min, 5:95 V/V at 12.0 min, 95:5 V/V at 12.1 min, 95:5 V/V at 15.0 min; flow rate, 0.40 mL/min; temperature, 40°C ; injection volume: $5^{\circ}\mu\text{L}$. The effluent was alternatively connected to an ESI-triple quadrupole-linear ion trap (Q TRAP)-MS system. Linear ion trap (LIT) and triple quadrupole (QQQ) scans were acquired on an API 4500 Q TRAP LC/MS/MS system, equipped with an ESI Turbo Ion-Spray interface, analyst software (AB Sciex, version 1.6) was used for instrument control, data acquisition, and subsequent quantification. A combination of both ionization modes (positive and negative) in MS full scan mode was applied for the molecular mass determination of the compounds in samples of peanut testa. The MS conditions for positive ionization modes optimized are as follows: the column oven was set at 35°C ; curtain gas, 30°psi ; ion spray voltage, 5,500 V; temperature, 600°C ; ion source gas 1, 60°psi ; and ion source gas 2, 50°psi . The MS conditions for negative ionization modes are as follows: the column oven was set at 35°C ; curtain gas, 30°psi ; ion spray voltage, $-4,500\text{ V}$; temperature, 600°C ; ion source gas 1, 60°psi ; and ion source gas 2, 50°psi . QQQ scans were acquired during MRM experiments with collision gas (nitrogen) set to 5°psi . DP and CE for individual MRM transitions were done with further DP and CE optimization. A specific set of MRM transitions was monitored for each period according to the metabolites eluted within the period. The MRM for each cultivar was performed in triplicate. Three technical replicates per sample were also analyzed in each series to assess technical reproducibility (Quality Controls: QCs). Through the overlapping display analysis of the total ions current (TIC) of MS detection and

analysis of quality controls (QCs), the average variation in metabolite abundance within the QCs (RSD) was < 20%. It shows that the signal stability of LC-MS/MS in different time periods is good.

Statistical analysis

All experiments were performed in triplicates and the results were presented as mean \pm standard deviation. Student's *t*-test and fold change analysis were used to compare the differential metabolites between the two groups. The significant differences [$p < 0.05$ and $|\log_2(\text{FC})| > 1$] among the mean values of different samples were analyzed by performing the Duncan test using IBM SPSS Statistics version 25 (IBM Corporation, New York, NY, USA). Correlation analysis used Pearson correlation coefficient, which measured the degree of linear correlation between two quantitative variables. The KEGG ID of different metabolites was used for pathway enrichment analysis to obtain the enrichment results of metabolic pathways.

Results and discussion

Phenotypic and extractive differences among the four peanut germplasms

The four peanut cultivars including Yuhua29 (YH29), Zhonghua12 (ZH12), Weihua10 (WH10), and Kainongbai (KNB) showed purple, red, pink, and white, testa colors respectively (Figure 1). The variation of metabolic components

in these four peanut genotypes with different testa colors was investigated using targeted metabolomics analysis. The target metabolites were detected qualitatively and quantitatively by the method of UPLC-ESI-MS/MS (Ultra performance liquid chromatography-electrospray ionization-tandem mass spectrometry). Clear quantitative and qualitative differences were revealed with a first inspection of the raw data plotted as TIC (Total ions current) and base peak chromatograms, respectively (Supplementary Figure 1A). Raw data was processed in a target manner based on the dedicated analysis package which included absolute quantitative data of 130 phenolic substances in 13 categories, namely anthocyanidins, flavones, flavonols, flavanones, isoflavones, coumarins, dihydrochalcones, phenylpropanoids, proanthocyanidins, benzoic acid derivatives, catechin derivatives, stilbenes, and terpenoids. Most of the above substances were belong to flavonoids, and the rest were polyphenols closely related to flavonoids. A total of 85 metabolites were identified in four peanuts (Supplementary Table 1). These identified metabolites belonged to different subgroups, including 9 phenylpropanoids, 13 benzoic acid derivatives, 7 catechin derivatives, 3 stilbenes, 2 dihydrochalcones, 5 anthocyanidins, 5 flavones, 16 flavonols, 9 flavanones, 2 terpenoids, 4 coumarins, 6 isoflavones, and 4 proanthocyanidins.

Principal Component Analysis (PCA) confirmed that the biological replicates of each sample were well-correlated. Moreover, there were obvious differences between the four testa with different colors, which also demonstrated the different metabolite composition among the four samples (Figure 2A). We found that 46 metabolites were present in all four samples, and 2, 5, 2, and 3 metabolites were specific in YH29, ZH12, WH10, and KNB, respectively (Figure 2B).



FIGURE 1

Phenotypic differences in the four testa of peanut. Differences of complete kernels, peeled kernels, and testa color phenotypes among four varieties on the period of DAP 70.

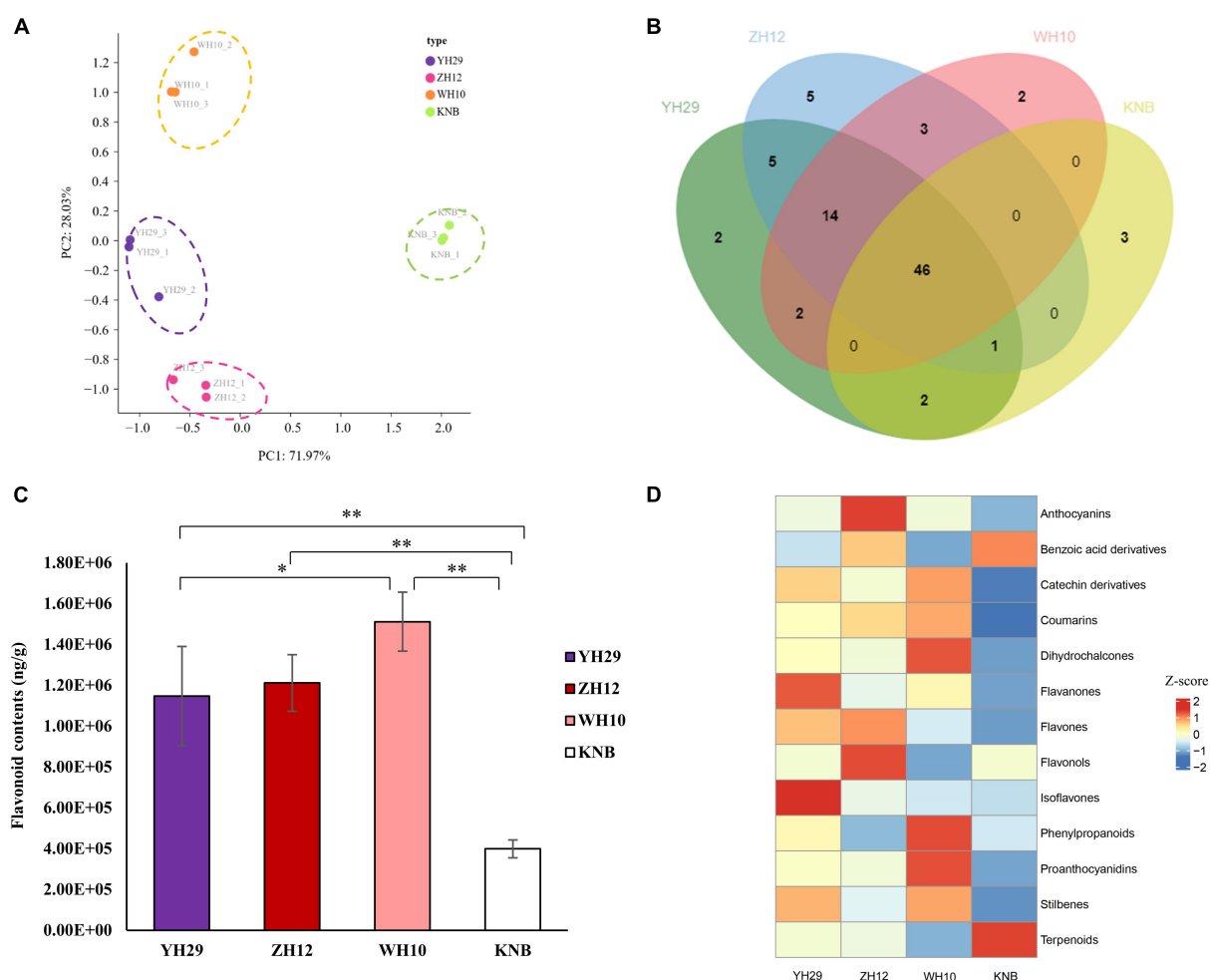


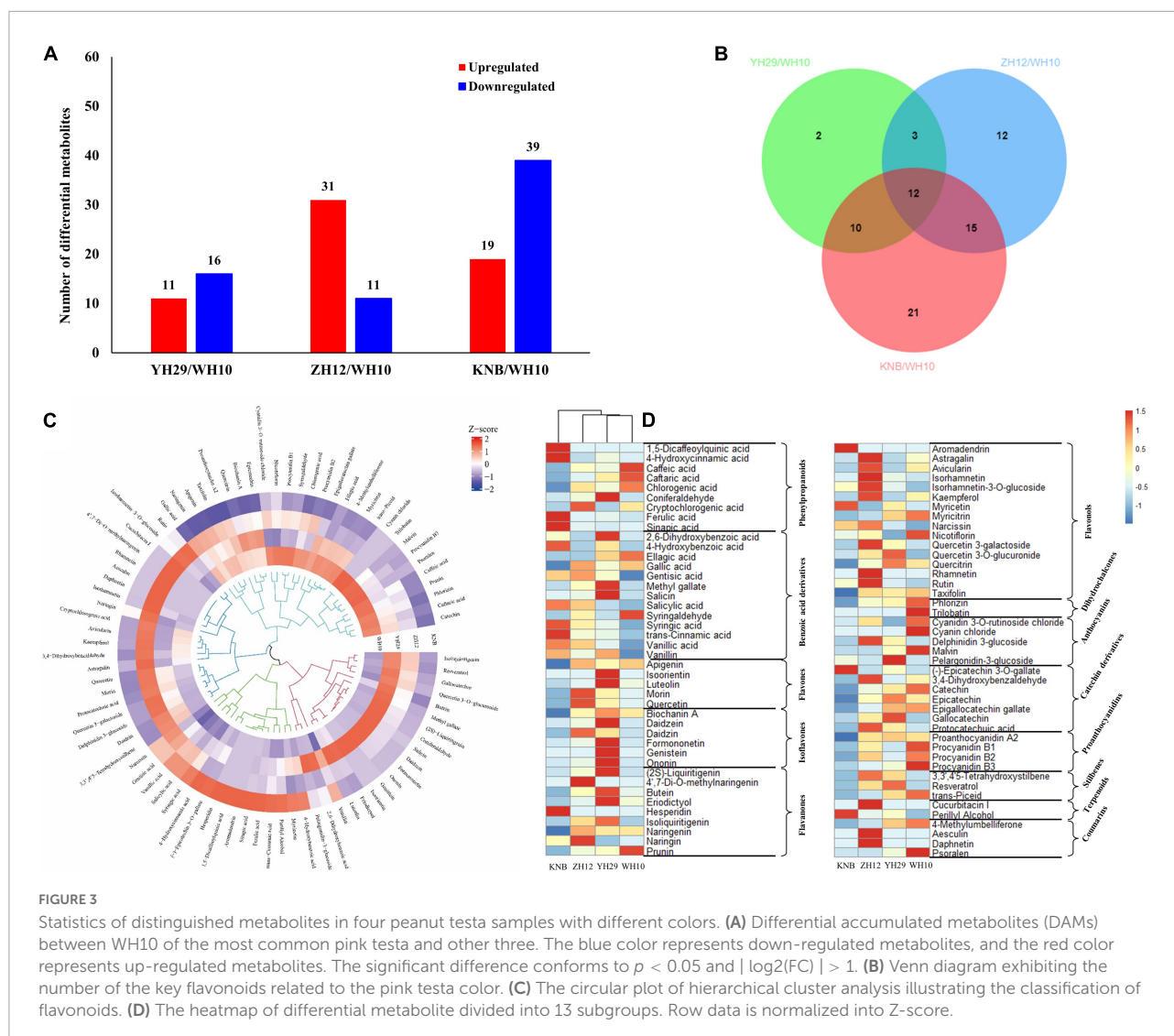
FIGURE 2

Extractive composition differences in type and content among the four samples. (A) Principal component analysis (PCA) score plots for metabolites in four types of peanut testa based on LC-MS/MS. (B) Distribution of the 85 flavonoids in the different colored testa. (C) The total content of flavonoids detected in the four samples. (D) The comparison of the relative contents of different type of flavonoids in the four samples. Row data is normalized into Z-score.

We found that the total content of flavonoids was highest in the pink testa of WH10, and lowest in the white testa of KNB (Figure 2C). The absolute content of various types of flavonoids revealed that proanthocyanidins, flavourols, catechin derivatives, phylpropanoids, and anthocyanins were the main flavonoids in peanut testa, while proanthocyanidins contributed more than half of the total flavonoids in the pink which contained the most flavonoids in four samples (Supplementary Figure 1B). The comparison of the relative contents of different groups of flavonoids in the four samples showed that the isoflavones and flavanones were abundant in YH29, anthocyanins, and flavourols were abundant in ZH12, phenopropionoids, proanthocyanidins, and dihydrochalcones were abundant in WH10, and terpenoids and benzoic acid derivatives were abundant in KNB, respectively (Figure 2D).

Distinguished metabolites in four peanut testa samples with different colors

In total, 78 differentially accumulated metabolites (DAMs) were identified across four testa colors (Supplementary Figure 2). Among them, 64 DAMs were differentially accumulated between the red (ZH12) and white (KNB) peanut testa colors, including 52 up-regulated and 12 down-regulated flavonoids (Supplementary Figure 3A). However, only 27 differentially expressed flavonoids were identified between purple (YH29) and pink (WH10) peanut testa colors (Figure 3A). We also observed that there was more kind of flavonoids in red testa peanut (ZH12) which had a total of 74 kinds, while there were only 52 kinds of flavonoids were identified in white (KNB) testa.



To find the key flavonoids in the formation of pink testa, Venn analysis was performed among the differential metabolites between YH29 and WH10, ZH12 and WH10, and KNB, and WH10 (Figure 3B). The result revealed that there were 12 overlapping significantly differential flavonoids among the three pairwise comparisons. Similarly, we compared the YH29, ZH12, and KNB against the other testa and detected 10, 15, and 30 overlapping differential metabolites, respectively (Supplementary Figure 3). In total, 49 overlapping differential flavonoids were obtained (Table 1). These overlapping flavonoids may play an important role in the development of different colors of peanut testa.

In order to visualize the accumulation patterns of metabolites in peanut testa with different colors, multivariate data analysis was performed (Figure 3C). Many specific metabolites were identified in different peanut testa colors, suggesting that the testa tissues might different metabolic

pathways synthesizing different pigments. In addition, we also found that each testa had higher accumulation of specific metabolites as compare to other testa colors, implying different metabolic directions of flavonoids existing in differently colored testa of peanuts (Figure 3D). For instance, the phenylpropanoids and benzoic acid derivatives content were higher in white testa (KNB) than other three peanut testa colors. Moreover, the content of secondary metabolites in white testa of KNB after entering in the flavonoid pathway was significantly lower than that of the other samples. Some of the ultimate products like anthocyanidins and proanthocyanidins were not detected in KNB, indicating that the metabolites of white testa of KNB did not enter in the flavonoid production after phenylalanine biosynthesis pathway which is primary stage of flavonoid production.

Moreover, we found that the red testa of ZH12 had a higher content of flavonols. The purple testa of YH29 had a high

TABLE 1 Key 49 significantly differential metabolites between the different colored samples.

Class	Compounds	Samples	KEGG ID
Anthocyanins	Cyanidin 3-O-rutinoside chloride	YH29 KNB	–
Anthocyanins	Cyanin chloride	WH10	–
Anthocyanins	Malvin	YH29 WH10	–
Anthocyanins	Pelargonidin-3-glucoside	YH29 KNB	C12137
Benzoic acid derivatives	Gentisic acid	WH10	C00628
Benzoic acid derivatives	Methyl gallate	YH29 ZH12	–
Benzoic acid derivatives	Salicin	YH29	C01451
Catechin derivatives	3,4-Dihydroxybenzaldehyde	WH10 YH29 ZH12 KNB	C16700
Catechin derivatives	Catechin	KNB	C06562
Catechin derivatives	Epicatechin	KNB	C09727
Catechin derivatives	Protocatechuic acid	WH10 KNB	C00230
Coumarins	4-Methylumbelliferone	ZH12 KNB	C03081
Coumarins	Aesculin	ZH12	C09264
Coumarins	Daphnetin	ZH12	C03093
Dihydrochalcones	Phlorizin	KNB	C01604
Dihydrochalcones	Trilobatin	WH10	–
Flavanones	4',7-Di-O-methylnaringenin	ZH12	–
Flavanones	Eriodictyol	KNB	C05631
Flavanones	Hesperidin	KNB	C09755
Flavanones	Naringenin	KNB	C00509
Flavanones	Prunin	WH10 KNB	–
Flavones	Apigenin	KNB	C01477
Flavones	Morin	KNB	C10105
Flavones	Quercetin	KNB	C00389
Flavonols	Aromadendrin	KNB	C00974
Flavonols	Astragalin	ZH12	C12249
Flavonols	Isorhamnetin	ZH12	C10084
Flavonols	Isorhamnetin-3-O-glucoside	ZH12 WH10	–
Flavonols	Kaempferol	ZH12 WH10	C05903
Flavonols	Nicotiflorin	YH29	–
Flavonols	Quercetin 3-galactoside	ZH12	C10073
Flavonols	Quercetin 3-O-glucuronide	WH10 KNB	–
Flavonols	Rhamnetin	ZH12	–
Flavonols	Rutin	ZH12	C05625
Flavonols	Taxifolin	KNB	–
Isoflavones	Genistein	YH29	C06563
Phenylpropanoids	1,5-Dicaffeoylquinic acid	KNB	C10445
Phenylpropanoids	4-Hydroxycinnamic acid	KNB	C00811
Phenylpropanoids	Caffeic acid	WH10 KNB	C01481
Phenylpropanoids	Caftaric acid	KNB	–
Phenylpropanoids	Chlorogenic acid	KNB	C00852
Phenylpropanoids	Ferulic acid	KNB	C01494
Phenylpropanoids	Sinapic acid	KNB	C00482
Proanthocyanidins	Proanthocyanidin A2	KNB	C10237
Proanthocyanidins	Procyanidin B1	YH29 KNB	–
Proanthocyanidins	Procyanidin B2	YH29 KNB	C17639
Proanthocyanidins	Procyanidin B3	WH10 KNB	–
Stilbenes	trans-Piceid	ZH12 KNB	C10275
Terpenoids	Cucurbitacin I	ZH12	C08800

level of flavonoids, flavanones and isoflavones, while pink testa of WH10 contained more downstream metabolites, such as dihydrochalcones, catechin derivatives, and proanthocyanidins (Figure 3D). Recently, the proanthocyanidins extracted from peanut testa were identified as a novel allosteric AKT inhibitor with potent anti-tumor efficacy beyond its antioxidant and anti-inflammatory properties (Liu et al., 2020). Here, we also observed that the total amount and types of proanthocyanidins which are terminal metabolites in flavonoid biosynthesis pathway were significantly higher in pink testa as compare to other peanut testa colors.

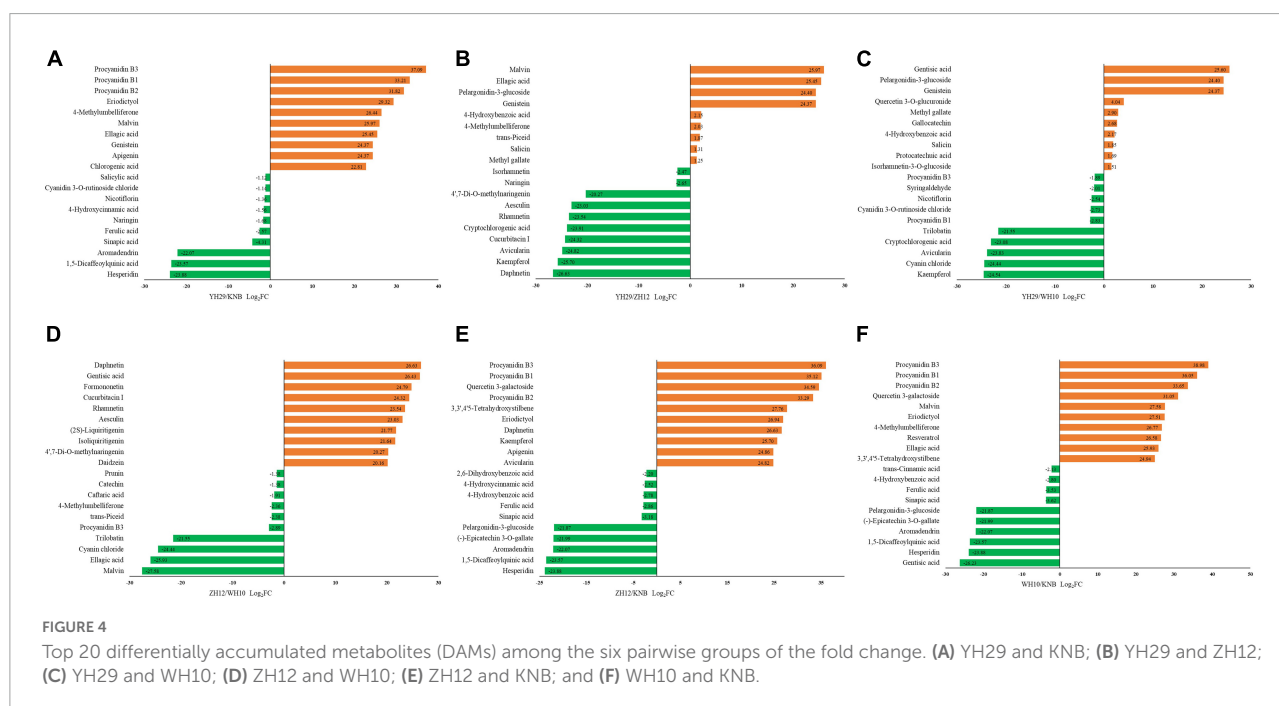
To better understand the variation of metabolic components between different peanut testa colors, we performed a pairwise comparison and identified top 20 various significant metabolites (Figure 4). The procyanidins content in pink testa (WH10) was significantly higher as compare to other testa colors. The purple testa (YH29) consistently showed higher anthocyanidin content. The red testa (ZH12) contained more coumarins and flavonols, such as daphnetin, aesculetin, kaempferol and avicularin. Meanwhile, phenylpropanoid, and benzoic acid derivatives were higher in white testa (KNB), for example 4-hydroxybenzoic acid, sinapic acid, 4-hydroxycinnamic acid, and ferulic acid.

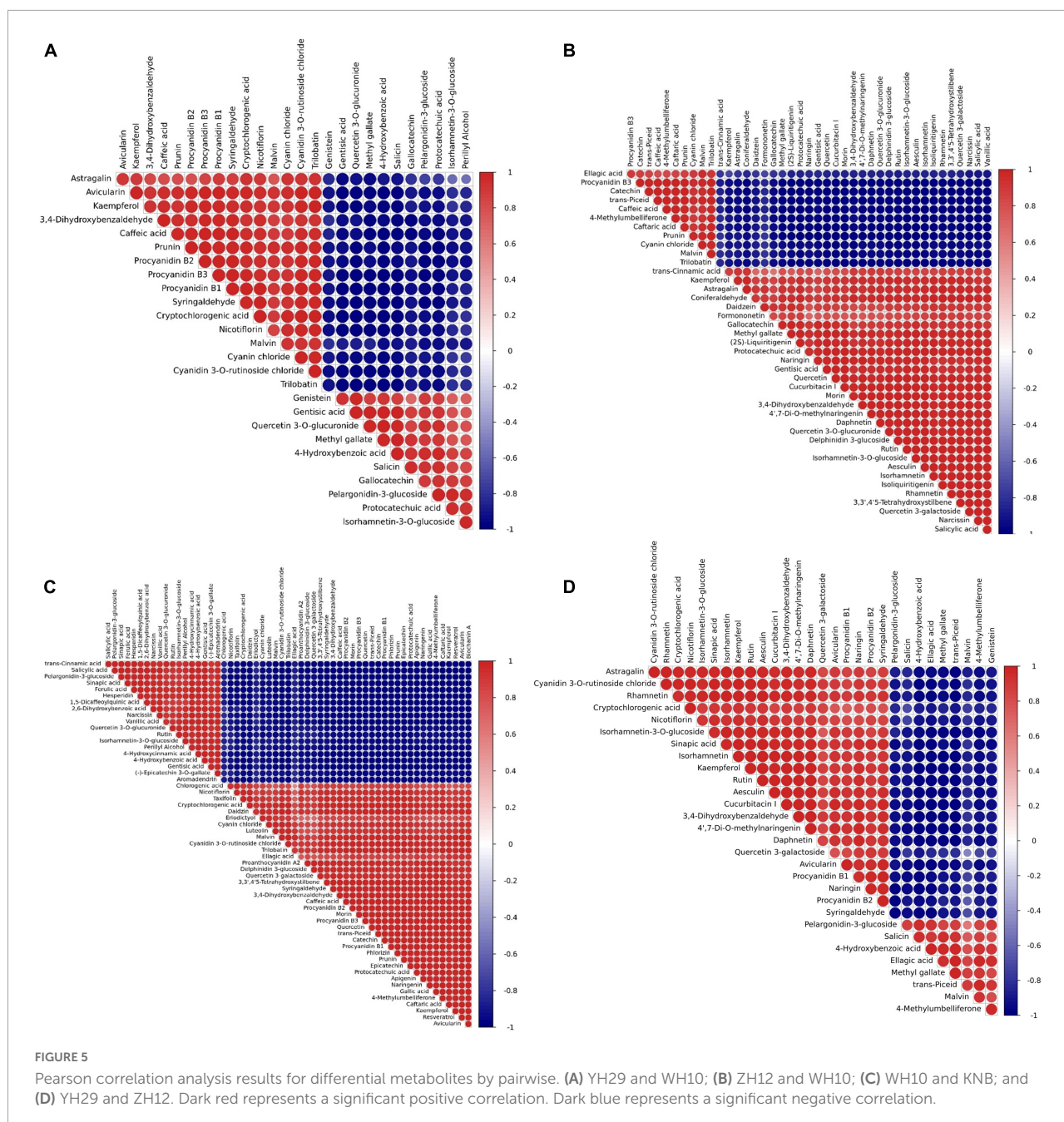
An increasing amount of evidence has showed that many flavonoids had biological activities, and the functions of flavonoids in nutrition and health care have been excavated (Supplementary Table 2). For example, the hydroxyl and carbonyl groups in the structure of most of the flavonoids could avoid oxidative damage by preventing the production of free radicals and scavenging their antioxidant activity

(Rice-evans et al., 1995; Burda and Oleszek, 2001). Along with the antioxidant function, some flavonoids had anticancer (Imran et al., 2019), antitumor (Selvendiran et al., 2006), anti-inflammatory (Lee et al., 2011), antibacterial (Comalada et al., 2005), analgesic (Khan, 2017), antiallergic (Baghel et al., 2012), and many other health benefits. In addition, some compounds could also be used as nutritional additives, spices or essence (Raskin, 1992; Suresh et al., 2009). We can use these findings to deploy targeted and precise breeding of peanuts, and develop peanut varieties with high-quality, high content of target metabolites to develop the peanuts as functional foods.

Correlation analysis of differential metabolites in peanut testa

Since there are diverse metabolites present in various colors of peanut testa, it is an interesting to analyze the association between different metabolic compounds. Pearson correlation analysis showed significant ($p < 0.05$) pairwise correlation based on 26 flavonoids between purple (YH29)/pink (WH10) testa colors (Figure 5A). While, pairwise correlation based on 41 flavonoids between red testa (ZH12)/pink testa (WH10), and 57 flavonoids between pink testa (WH10)/white (KNB) (Figures 5B,C). The performance of some flavonoids in different correlation comparison groups was consistent. For instance, proanthocyanidins B1, B2, and B3 in pairs were consistently showed positive correlation. Proanthocyanidins were positively correlated with cyanidin 3-O-rutinoside and

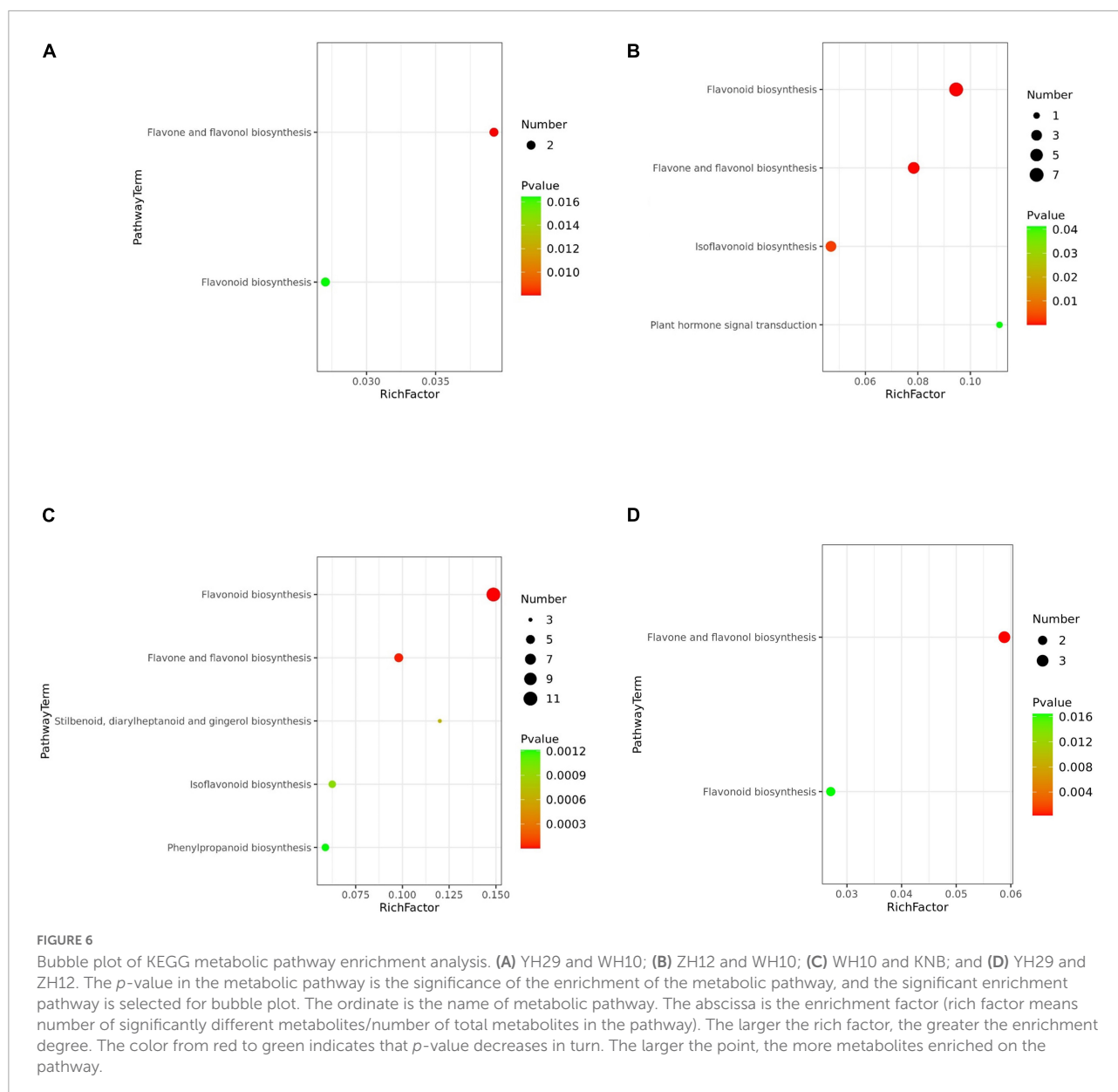




malvin, but negatively correlated with pelargonidin-3-glucoside (Figure 5D). Interestingly, these are all anthocyanidins. There were also some flavonoids with different correlations in different groups. For instance, proanthocyanidins were positively correlated with kaempferol and astragalol in YH29/WH10, WH10/KNB, YH29/ZH12, and ZH12/KNB, while negatively correlated in ZH12/WH10 (Supplementary Figure 4). These results indicated that the flavonoid metabolic pathway is basically same in different testa color varieties, however some changes might occur due to individual gene variations in the pathway.

Enrichment analysis of metabolic pathways and flow direction of metabolites

In the pink testa peanut WH10, the pathways of “Flavone and flavonol biosynthesis” and “Flavonoid biosynthesis” were highly enriched (Figure 6A). Flavonoid biosynthesis, flavone and flavonol biosynthesis were significantly differential metabolites between red (ZH12) and pink (WH10) were mostly involved in flavonoid biosynthesis, (Figure 6B). The differential metabolites between white (KNB) and pink



(WH10) were involved flavonoid biosynthesis, flavone and flavonol biosynthesis (Figure 6C). Our data indicated that the differential metabolites between YH29 and ZH12 were enriched mainly in flavone and flavonol biosynthesis (Figure 6D). The KEGG between YH29/KNB and ZH12/KNB enriched mainly in flavonoid biosynthesis, flavone and flavonol biosynthesis, isoflavonoid biosynthesis, and phenylpropanoid biosynthesis (Supplementary Figure 5).

A schematic diagram was deduced to illustrate the selected differential metabolites among the six pairwise comparisons in the flavonoid biosynthesis pathway (Figure 7). Compared with the pink (WH10), the flavonoids in the testa of red (ZH12) obviously flowed to the isoflavone, flavonol and anthocyanin synthesis pathways. In contrast, the

content of proanthocyanidin and catechin, which was the proanthocyanidin synthetic monomers, in WH10 is significantly higher than that in ZH12. Therefore, we concluded that there are several "key nodes" or "switches" in the pink and red testa peanut varieties, which controlled the accumulation and flow direction of flavonoids which results in significant differences in flavonoids metabolites in testa of red and pink color. We speculated that the most likely "key nodes" between red (ZH12) and pink (WH10) were the branches from the main flavonoid biosynthesis pathway to the isoflavone pathway, to the flavonol pathway and to the proanthocyanidin pathway.

The obviously different flavonoid accumulation patterns were also observed between pink (WH10) and white (KNB) testa colors (Figure 7). The compounds in the middle

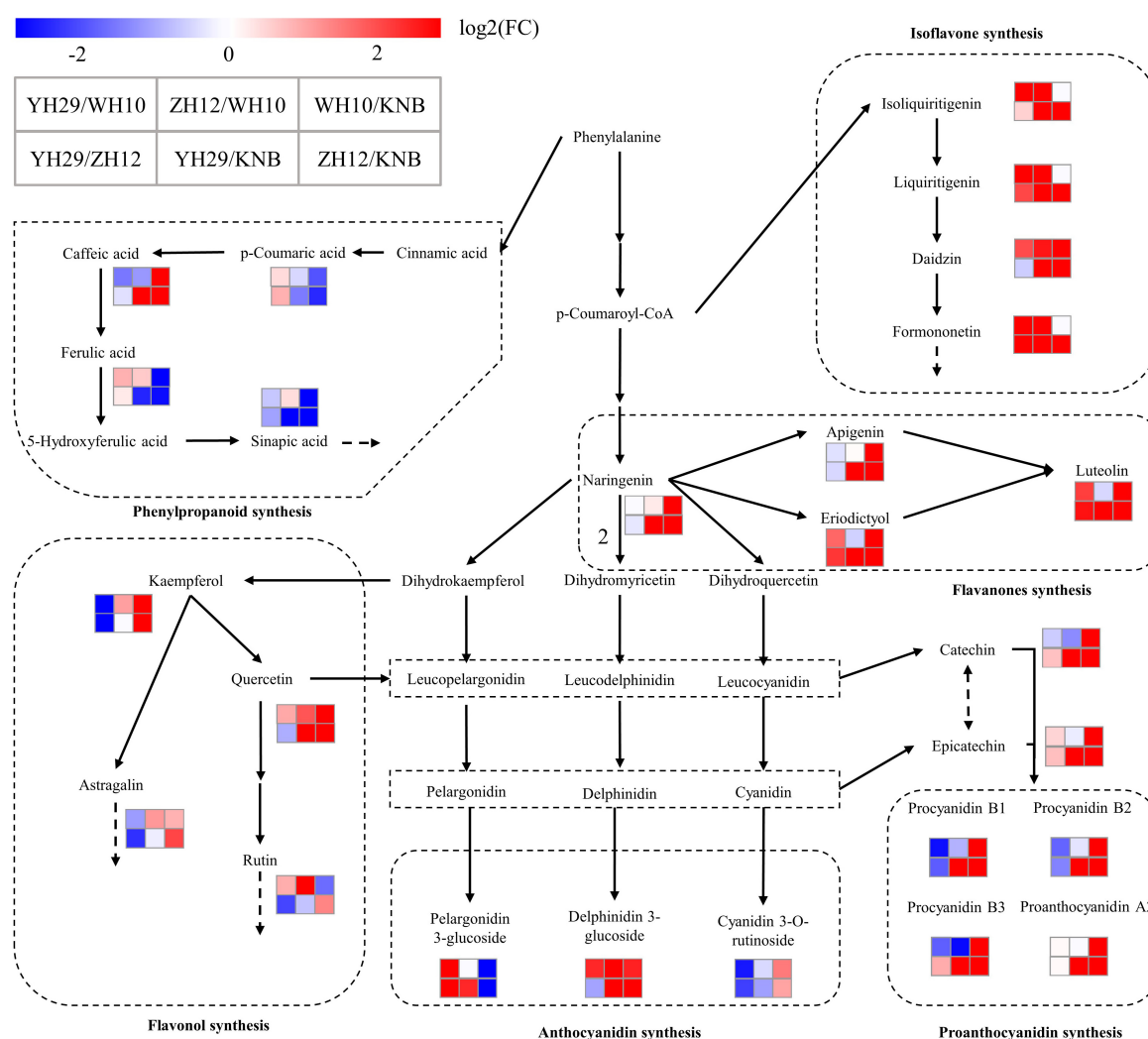


FIGURE 7

Schematic diagram of the metabolic pathway and relative content of metabolites. (A) ZH12 and WH10; (B) WH10 and KNB. The absolute content is shown in the small histogram next to the compound using the average of all values. The metabolic pathway was modified based on the KEGG database (<http://www.genome.jp/kegg/>).

and downstream of flavonoid synthesis pathway accumulated significantly in pink (WH10) testa, while in white (KNB) testa, the derivative content of phenylalanine, which were the flavonoid pathway starting substance, was much higher than that in pink (WH10) testa. The content of naringenin in pink peanut was also much higher compare to white testa peanut, we speculated that this "key nodes" should be before the synthesis of naringenin, or even at the initiation of phenylalanine biosynthesis.

In previous studies, some key genes have been found to regulate anthocyanin accumulation in peanut testa. The candidate gene controlling peanut purple testa was considered to be one of two MYB transcription factors (Zhao et al., 2019; Huang et al., 2020). Two independently inherited genes might be controlling peanut red testa, One encoded bHLH transcription

factor and the other encoded anthocyanin reductase (Chen et al., 2021; Zhang et al., 2022). By analyzing the pink and red testa of peanut by transcriptome and metabolome, some CHS, F3'H, DFR, MYB, bHLH, and WD40 genes may be the key formation and regulatory genes controlling the formation of pink and red testa in peanut (Xue et al., 2021). FLS, DFR, and WD40 transcription factor gene might be the key responsible for the white testa phenotype (Wan et al., 2020). Combined with our research, these key genes may be the structural genes at the node, or regulatory factors (for example, some transcription factors) acting at the node, which the changes of these might be the root cause behind the difference in metabolite accumulation in various peanut testa colors. Our study is helpful to determine the function of these genes and the molecular mechanism of regulation.

Conclusion

In this study, we performed accurately targeted metabolite profiling by UPLC-MS/MS of four types peanut testa colors to evaluate the metabolite differences. We observed that the different accumulation patterns of flavonoids in peanut testa of different colors are related to the flow of flavonoid metabolites across the pathways. The different effects of structural genes or regulatory factors near nodes in different testa color peanuts might lead to the final flavonoid metabolome differences. Moreover, we also discussed the nutritional and health care effects of flavonoids, and the prospect of breeding of peanut varieties high flavonoids content.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding authors.

Author contributions

YZ and CZ conceived the project. KZ, JM, SG, LH, HX, NL, JP, RT, and HH carried out the experiments and analyzed the experimental results. KZ and JM wrote the original draft. CZ, XW, YZ, and SG helped to review and revise the original draft. All authors have read and approved the final manuscript.

Funding

This research was supported by National Natural Science Foundation of China (32072090 and 31861143009), Key Research and Development Project of Shandong Province (2020LZGC001 and 2021LZGC025), Shandong Province Natural Science Foundation (ZR2020MC104 and ZR2020MC105), Famous subjects construction project of Shandong Agricultural and Engineering University, Agricultural scientific and technological innovation project of Shandong Academy of Agricultural Sciences, and Taishan Scholar Project of Shandong Province (ts20190964).

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.992124/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Extractive composition differences in type and content among the four samples. (A) The total ions current (TIC) of the samples. The abscissa is the retention time (RT) of the metabolite detection, and the ordinate is the ionic strength (count per second, CPS) of the ion detection. (B) The absolute content of various types of flavonoids.

SUPPLEMENTARY FIGURE 2

Volcano plots of differential metabolites in the pairwise comparison between: (A) YH29 and KNB; (B) YH29 and ZH12; (C) YH29 and WH10; (D) ZH12 and WH10; (E) ZH12 and KNB; and (F) WH10 and KNB. Each point in the volcano map represents a metabolite, the abscissa represents the logarithmic value of the difference of relative content of a certain metabolite in the two samples, the ordinate represents the VIP value. The greater the absolute value of the abscissa, the greater the multiple difference in the expression level between the two samples; the greater the ordinate value, the more significant the differential expression, and the more reliable the differentially expressed metabolites screened. In the figure, the blue dots represent down-regulated differentially metabolites, the red dots represent up-regulated differentially metabolites, and gray represents detected but not significantly different metabolites.

SUPPLEMENTARY FIGURE 3

Statistics of distinguished metabolites in four peanut testa samples with different colors. (A) Differential accumulated metabolites (DAMs) among YH29/KNB, ZH12/KNB, and YH29/ZH12. The blue color represents down-regulated metabolites, and the red color represents up-regulated metabolites. The significant difference conforms to $p < 0.05$ and $|\log_2(FC)| > 1$. Venn diagram exhibiting the number of the key flavonoids related to (B) YH29; (C) ZH12; and (D) KNB.

SUPPLEMENTARY FIGURE 4

Pearson correlation analysis results for differential metabolites by pairwise. (A) YH29 and KNB; (B) ZH12 and KNB. Dark red represents a significant positive correlation. Dark blue represents a significant negative correlation.

SUPPLEMENTARY FIGURE 5

Bubble plot of KEGG metabolic pathway enrichment analysis. (A) YH29 and KNB; (B) ZH12 and KNB. The p -value in the metabolic pathway is the significance of the enrichment of the metabolic pathway, and the significant enrichment pathway is selected for bubble plot. The ordinate is the name of metabolic pathway. The abscissa is the enrichment factor (rich factor means number of significantly different metabolites/number of total metabolites in the pathway). The larger the rich factor, the greater the enrichment degree. The color from red to green indicates that p -value decreases in turn. The larger the point, the more metabolites enriched on the pathway.

References

- Baghel, S. S., Shrivastava, N., Baghel, R. S., Agrawal, P., and Rajput, S. (2012). A review of quercetin: Antioxidant and anticancer properties. *World J. Pharm. Pharmaceutical Sci.* 1, 146–160.
- Bradshaw, H., and Schemske, D. W. (2003). Allele substitution at a flower colour locus produces a pollinator shift in monkeyflowers. *Nature* 426, 176–178. doi: 10.1038/nature02106
- Burda, S., and Oleszek, W. (2001). Antioxidant and antiradical activities of flavonoids. *J. Agric. Food Chem.* 49, 2774–2779. doi: 10.1021/jf001413m
- Cavauiuolo, M., Cocetta, G., and Ferrante, A. (2013). The antioxidants changes in ornamental flowers during development and senescence. *Antioxidants* 2, 132–155. doi: 10.3390/antiox2030132
- Chen, H., Chen, X., Xu, R., Liu, W., Liu, N., Huang, L., et al. (2021). Fine-mapping and gene candidate analysis for AhRt1, a major dominant locus responsible for testa color in cultivated peanut. *Theor. Appl. Genet.* 134, 3721–3730. doi: 10.1007/s00122-021-03924-w
- Christman, L. M., Dean, L. L., Bueno Almeida, C., and Weissburg, J. R. (2018). Acceptability of peanut skins as a natural antioxidant in flavored coated peanuts. *J. Food Sci.* 83, 2571–2577. doi: 10.1111/1750-3841.14323
- Comalada, M., Camuesco, D., Sierra, S., Ballester, I., Xaus, J., Gálvez, J., et al. (2005). In vivo quercitrin anti-inflammatory effect involves release of quercetin, which inhibits inflammation through down-regulation of the NF- κ B pathway. *Eur. J. Immunol.* 35, 584–592. doi: 10.1002/eji.200425778
- Constanza, K. E., White, B. L., Davis, J. P., Sanders, T. H., and Dean, L. L. (2012). Value-added processing of peanut skins: Antioxidant capacity, total phenolics, and procyanidin content of spray-dried extracts. *J. Agric. Food Chem.* 60, 10776–10783. doi: 10.1021/jf3035258
- Crozier, A., Jaganath, I. B., and Clifford, M. N. (2009). Dietary phenolics: Chemistry, bioavailability and effects on health. *Nat. Prod. Rep.* 26, 1001–1043. doi: 10.1039/b802662a
- Dong, N. Q., and Lin, H. X. (2021). Contribution of phenylpropanoid metabolism to plant development and plant-environment interactions. *Integr. Plant Biol.* 63, 180–209. doi: 10.1111/jipb.13054
- Dong, T., Han, R., Yu, J., Zhu, M., Zhang, Y., Gong, Y., et al. (2019). Anthocyanins accumulation and molecular analysis of correlated genes by metabolome and transcriptome in green and purple asparagus (*Asparagus officinalis* L.). *Food Chem.* 271, 18–28. doi: 10.1016/j.foodchem.2018.07.120
- Francisco, M. L. L. D., and Resurreccion, A. (2009). Total phenolics and antioxidant capacity of heat-treated peanut skins. *J. Food Compos. Anal.* 22, 16–24. doi: 10.1016/j.jfca.2008.05.012
- 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
- Hathorn, C. S., and Sanders, T. H. (2012). Flavor and antioxidant capacity of peanut paste and peanut butter supplemented with peanut skins. *J. Food Sci.* 77, S407–S411. doi: 10.1111/j.1750-3841.2012.02953.x
- Huang, L., Liu, X., Pandey, M. K., Ren, X., Chen, H., Xue, X., et al. (2020). Genome-wide expression quantitative trait locus analysis in a recombinant inbred line population for trait dissection in peanut. *Plant Biotechnol. J.* 18, 779–790. doi: 10.1111/pbi.13246
- Imran, M., Salehi, B., Sharifi-Rad, J., Aslam Gondal, T., Saeed, F., Imran, A., et al. (2019). Kaempferol: A key emphasis to its anticancer potential. *Molecules* 24:2277. doi: 10.3390/molecules24122277
- Iwashina, T. (2003). Flavonoid function and activity to plants and other organisms. *Biol. Sci. Space* 17, 24–44. doi: 10.2187/bss.17.24
- Karadag, A., Ozcelik, B., and Saner, S. (2009). Review of methods to determine antioxidant capacities. *Food Anal. Methods* 2, 41–60. doi: 10.1007/s12161-008-9067-7
- Karchesy, J. J., and Hemingway, R. W. (1986). Condensed tannins: (4. beta. fvdarw. 8; 2. beta. fvdarw. O. fvdarw. 7)-linked procyanidins in *Arachis hypogaea* L. *J. Agric. Food Chem.* 34, 966–970. doi: 10.1021/jf00072a009
- Khan, A. S. (2017). “Antipyretic and analgesic activities of some economically important woody plants,” in *Medicinally important trees*, doi: 10.1007/978-3-319-56777-8_7 (Berlin: Springer), 159–185.
- Kuang, Q., Yu, Y., Attree, R., and Xu, B. (2017). A comparative study on anthocyanin, saponin, and oil profiles of black and red seed coat peanut (*Arachis hypogaea*) grown in China. *Int. J. Food Properties* 20(sup1), S131–S140. doi: 10.1080/10942912.2017.1291676
- Lee, H. B., Kim, E. K., Park, S. J., Bang, S. G., Kim, T. G., and Chung, D. W. (2011). Isolation and anti-inflammatory effect of astragalol synthesized by enzymatic hydrolysis of tea seed extract. *J. Sci. Food Agric.* 91, 2315–2321. doi: 10.1002/jsfa.4457
- Liu, G., Shi, A., Wang, N., Li, M., He, X., Yin, C., et al. (2020). Polyphenolic Proanthocyanidin-B2 suppresses proliferation of liver cancer cells and hepatocellular carcinogenesis through directly binding and inhibiting AKT activity. *Redox Biol.* 37:101701. doi: 10.1016/j.redox.2020.101701
- 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
- Nabavi, S. M., Samec, D., Tomczyk, M., Milella, L., Russo, D., Habtemariam, S., et al. (2020). Flavonoid biosynthetic pathways in plants: Versatile targets for metabolic engineering. *Biotechnol. Adv.* 38:107316. doi: 10.1016/j.biotechadv.2018.11.005
- Nepote, V., Grosso, N., and Guzman, C. A. (2007). Extraction of antioxidant components from peanut skins. *Grasas Aceites* 53, 391–395. doi: 10.3989/gya.2002.v53.i4.335
- Noda, N., Yoshioka, S., Kishimoto, S., Nakayama, M., Douzono, M., Tanaka, Y., et al. (2017). Generation of blue chrysanthemums by anthocyanin B-ring hydroxylation and glucosylation and its coloration mechanism. *Sci. Adv.* 3:e1602785. doi: 10.1126/sciadv.1602785
- Oliveira, T. K. B. d., Gomes, J. P., Silva Júnior, P. R. d., Lima, A. R. N., Jordão, A. J. J. M. d. L., Ramos, K. R. d. L. P., et al. (2021). Morphological, biochemical and histological effects of aqueous extracts of peanut (*Arachis hypogaea*) on swiss mice in different diets. *Acta Cirúrgica Brasileira* 36:e360905. doi: 10.1590/ACB360905
- Osakabe, N., and Yamagishi, M. (2009). Procyanidins in Theobroma cacao reduce plasma cholesterol levels in high cholesterol-fed rats. *J. Clin. Biochem. Nutr.* 45, 131–136. doi: 10.3164/jcbs.07-34
- Park, Y. S., Jeon, M. H., Hwang, H. J., Park, M. R., Lee, S.-H., Kim, S. G., et al. (2011). Antioxidant activity and analysis of proanthocyanidins from pine (*Pinus densiflora*) needles. *Nutr. Res. Pract.* 5, 281–287. doi: 10.4162/nrp.2011.5.4.281
- Pinent, M., Blay, M., Bladé, M., Salvado, M., Arola, L., and Ardevol, A. (2004). Grape seed-derived procyanidins have an antihyperglycemic effect in streptozotocin-induced diabetic rats and insulinomimetic activity in insulin-sensitive cell lines. *Endocrinology* 145, 4985–4990. doi: 10.1210/en.2004-0764
- Pourcel, L., Routaboul, J.-M., Cheynier, V., Lepiniec, L., and Debeaujon, I. (2007). Flavonoid oxidation in plants: From biochemical properties to physiological functions. *Trends Plant Sci.* 12, 29–36. doi: 10.1016/j.tplants.2006.11.006
- Raskin, I. (1992). Role of salicylic acid in plants. *Annu. Rev. Plant Biol.* 43, 439–463. doi: 10.1146/annurev.pp.43.060192.002255
- Rice-evans, C. A., Miller, N. J., Bolwell, P. G., Bramley, P. M., and Pridham, J. B. (1995). The relative antioxidant activities of plant-derived polyphenolic flavonoids. *Free Radical Res.* 22, 375–383. doi: 10.3109/10715769509145649
- Sasaki, N., and Nakayama, T. (2015). Achievements and perspectives in biochemistry concerning anthocyanin modification for blue flower coloration. *Plant Cell Physiol.* 56, 28–40. doi: 10.1093/pcp/pcu097
- Selvendiran, K., Koga, H., Ueno, T., Yoshida, T., Maeyama, M., Torimura, T., et al. (2006). Luteolin promotes degradation in signal transducer and activator of transcription 3 in human hepatoma cells: An implication for the antitumor potential of flavonoids. *Cancer Res.* 66, 4826–4834. doi: 10.1158/0008-5472.CAN-05-4062
- Sirerol, J. A., Rodríguez, M. L., Mena, S., Asensi, M. A., Estrela, J. M., and Ortega, A. L. (2016). Role of natural stilbenes in the prevention of cancer. *Oxid. Med. Cell. Long.* 2016, 3128951. doi: 10.1155/2016/3128951
- Sun, C., Zhang, M., Dong, H., Liu, W., Guo, L., and Wang, X. (2020). A spatially-resolved approach to visualize the distribution and biosynthesis of flavones in

- Scutellaria baicalensis Georgi. *J. Pharmaceutical Biomed. Anal.* 179:113014. doi: 10.1016/j.jpba.2019.113014
- Suresh, D., Gurudutt, K., and Srinivasan, K. (2009). Degradation of bioactive spice compound: Curcumin during domestic cooking. *Eur. Food Res. Technol.* 228, 807–812. doi: 10.1007/s00217-008-0993-9
- Tanaka, Y., Brugliera, F., and Chandler, S. (2009). Recent progress of flower colour modification by biotechnology. *Int. J. Mol. Sci.* 10, 5350–5369. doi: 10.3390/ijms10125350
- Wan, L., Lei, Y., Yan, L., Liu, Y., Pandey, M. K., Wan, X., et al. (2020). Transcriptome and metabolome reveal redirection of flavonoids in a white testa peanut mutant. *BMC Plant Biol.* 20:161. doi: 10.1186/s12870-020-02383-7
- 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
- Xue, Q., Zhang, X., Yang, H., Li, H., Lv, Y., Zhang, K., et al. (2021). Transcriptome and metabolome analysis unveil anthocyanin metabolism in pink and red testa of peanut (*Arachis hypogaea* L.). *Int. J. Genom.* 2021:5883901. doi: 10.1155/2021/5883901
- Yu, J., Ahmedna, M., and Goktepe, I. (2005). Effects of processing methods and extraction solvents on concentration and antioxidant activity of peanut skin phenolics. *Food Chem.* 90, 199–206. doi: 10.1016/j.foodchem.2004.03.048
- Yu, J., Ahmedna, M., Goktepe, I., and Dai, J. (2006). Peanut skin procyanidins: Composition and antioxidant activities as affected by processing. *J. Food Compos. Anal.* 19, 364–371. doi: 10.1016/j.jfca.2005.08.003
- Zhang, K., Yuan, M., Xia, H., He, L., Ma, J., Wang, M., et al. (2022). BSA-seq and genetic mapping reveals AhRt2 as a candidate gene responsible for red testa of peanut. *Theor. Appl. Genet.* 135, 1529–1540. doi: 10.1007/s00122-022-04051-w
- Zhang, P., Du, H., Wang, J., Pu, Y., Yang, C., Yan, R., et al. (2020). Multiplex CRISPR/Cas9-mediated metabolic engineering increases soya bean isoflavone content and resistance to soya bean mosaic virus. *Plant Biotechnol. J.* 18, 1384–1395. doi: 10.1111/pbi.13302
- Zhao, Y., Ma, J., Li, M., Deng, L., Li, G., Xia, H., et al. (2019). Whole-genome resequencing-based QTL-seq identified AhTc1 gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. *Plant Biotechnol. J.* 18, 96–105. doi: 10.1111/pbi.13175



OPEN ACCESS

EDITED BY

Xia Xin,
Institute of Crop Sciences (CAAS),
China

REVIEWED BY

Suoyi Han,
Henan Academy of Agricultural
Sciences, China
Weihua Qiao,
Institute of Crop Sciences (CAAS),
China
Dongxin Huai,
Oil Crops Research Institute (CAAS),
China

*CORRESPONDENCE

Yan-an Guan
yguan65@163.com

SPECIALTY SECTION

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

RECEIVED 15 June 2022

ACCEPTED 16 August 2022

PUBLISHED 20 September 2022

CITATION

Li F-f, Niu J-h, Yu X, Kong Q-h,
Wang R-f, Qin L, Chen E-y, Yang Y-b,
Liu Z-y, Lang L-n, Zhang H-w,
Wang H-l and Guan Y-a (2022)
Isolation and identification of *SiCOL5*,
which is involved in photoperiod
response, based on the quantitative
trait locus mapping of *Setaria italica*.
Front. Plant Sci. 13:969604.
doi: 10.3389/fpls.2022.969604

COPYRIGHT

© 2022 Li, Niu, Yu, Kong, Wang, Qin,
Chen, Yang, Liu, Lang, Zhang, Wang
and Guan. 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(s)
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.

Isolation and identification of *SiCOL5*, which is involved in photoperiod response, based on the quantitative trait locus mapping of *Setaria italica*

Fei-fei Li¹, Jia-hong Niu², Xiao Yu², Qing-hua Kong²,
Run-feng Wang¹, Ling Qin¹, Er-ying Chen¹, Yan-bing Yang¹,
Zhen-yu Liu¹, Li-na Lang³, Hua-wen Zhang¹, Hai-lian Wang¹
and Yan-an Guan^{1*}

¹Featured Crops Engineering Laboratory of Shandong Province, National Engineering Research Center of Wheat and Maize, Shandong Technology Innovation Center of Wheat, Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, China, ²College of Life Science, Shandong Normal University, Jinan, China, ³Shandong Seed Administration Station, Jinan, China

Foxtail millet (*Setaria italica*) is a versatile grain and fodder crop grown in arid and semi-arid regions. It is an especially important crop for combating malnutrition in certain poverty-stricken areas of the world. Photoperiod sensitivity is a major constraint to the distribution and utilization of foxtail millet germplasm resources. Foxtail millet may be suitable as a model species for studying the photoperiod sensitivity of C₄ crops. However, the genetic basis of the photoperiod response of foxtail millet remains poorly studied. To detect the genetic basis of photoperiod sensitivity-related traits, a recombinant inbred line (RIL) population consisting of 313 lines derived from a cross between the spring-sown cultivar “Longgu 3” and the summer-sown cultivar “Canggu 3” was established. The RIL population was genotyped using whole-genome re-sequencing and was phenotyped in four environments. A high-density genetic linkage map was constructed with an average distance between adjacent markers of 0.69 cM. A total of 21 quantitative trait loci (QTLs) were identified by composite interval mapping, and 116 candidate genes were predicted according to gene annotations and variations between parents, among which three genes were considered important candidate genes by the integration and overall consideration of the results from gene annotation, SNP and indel analysis, *cis*-element analysis, and the expression pattern of different genes in different varieties, which have different photoperiod sensitivities. A putative candidate gene, *SiCOL5*, was isolated based on QTL mapping analysis. The expression of *SiCOL5* was sensitive to photoperiod and was regulated by biological rhythm-related genes. Function analysis

suggested that *SiCOL5* positively regulated flowering time. Yeast two-hybrid and bimolecular fluorescence complementation assays showed that *SiCOL5* was capable of interacting with *SiNF-YA1* in the nucleus.

KEYWORDS

foxtail millet, high-density linkage map, photoperiod sensitivity, QTL mapping, *SiCOL5*

Introduction

Foxtail millet (*Setaria italica*) is among the first domesticated grain crops and is a versatile grain and fodder crop cultivated in arid and semi-arid regions. It is an important crop for combating malnutrition in areas with chronic poverty. Foxtail millet is a short-day crop and is sensitive to photoperiod. Therefore, the plant has extremely limited ecological adaptability, and photoperiod sensitivity is a major constraint to the exchange of foxtail millet germplasm resources. It also accounts for the scarcity of broadly adaptable foxtail millet cultivars in commercial production, which limits further development and capacity to strengthen the competitiveness of the millet seed industry. Therefore, improvement of the adaptability of foxtail millet cultivars is important to expand and strengthen the millet seed industry (Jia and Diao, 2022).

Photoperiod is an important factor that regulates the flowering period of crops to ensure that flowering occurs at the correct time, which determines the regional adaptability of crops. Research on photoperiod regulation of flowering has mainly focused on model plants, such as *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza sativa*). *Arabidopsis* is a classic model plant, and its photoperiodic regulation of flowering has been studied in considerable detail, which can be summarized as follows: When there is an optical signal, *PHY* and *CRY* receive the optical signal, and through integration of the *ELF3* and *ZTL* genes, the signal is transmitted to the core oscillator, which then outputs the signal to *GI*, *FKF1*, *CO*, and other downstream genes. *CO* promotes flowering by inducing *FT*, *LFY*, and *SOC1* (Anwer et al., 2018; Sim et al., 2019). As a typical gramineous crop, research on rice provides an important reference for other gramineous crops. Many crucial genes in the photoperiodic control of flowering have been detected in rice by quantitative trait locus (QTL) mapping, such as *Hd3a*, *Hd1*, *Ehd1*, and *Ghd7* (Yano, 2000; Kazuyuki et al., 2004; Tsuji et al., 2008; Xue et al., 2009); two pathways promote flowering in a

short-day environment, namely, *OsGI-Hd1-Hd3a* and *OsGI-Ehd1-Hd3a*. Similarly, two pathways restrain flowering in a long-day environment, namely, *OsELF3-OsGI-Hd1-Hd3a* and *PhyA-Ghd7-Ehd1-Hd3a*, and one pathway promotes flowering in a long-day environment, namely, *OsMADS50-Ehd1-RFT1* (Lee et al., 2004). However, rice is a C_3 crop and is only distantly related to C_4 gramineous crops, such as corn (*Zea mays*), sorghum (*Sorghum bicolor*), sugarcane (*Saccharum* spp.), and foxtail millet.

Among C_4 crops, most studies of photoperiod sensitivity have focused on maize and sorghum. Several important genes associated with photoperiod, such as *SbEhd1*, *SbCO*, *SbPRR37*, and *SbGHD7*, have been cloned from sorghum (Murphy et al., 2011; Yang et al., 2014; Klein et al., 2015). Photoperiod response in maize has been studied extensively using QTL mapping (Chardon et al., 2004; Wang et al., 2008; Coles et al., 2010). The results of QTL mapping in maize indicate that the bin 10.04 region of chromosome 10 had the greatest additive effects and was the most stable QTL. The research reported that this maize chromosomal region includes at least three photoperiod-sensitive genes: *ZmCCT*, *ZmPIF3*, and *ZmCCA1*. In recent years, the functions of several major genes have been studied. In maize, *ZCN8* and *Conz1* are considered to be homologs of *FT* and *CO*, respectively, in *Arabidopsis* (Miller et al., 2008). The gene *ZmCCT* is a homolog of *Ghd7* in rice. Jin et al. (2018) speculated that *ZmCOL3* might inhibit flowering through positive regulation of *ZmCCT* expression or inhibition of the biological clock. Cheng et al. (2018) showed that *ZmCCT9* inhibited flowering by suppressing *ZCN8* expression. Su et al. (2018) speculated that *ZmNF-YA3* may form trimers with *CO*-like and *PPF1* and bind to the promoter of *ZmFT-like12* to promote its expression and thus accelerate the flowering transition. Thus, several photoperiod-sensitive genes have been identified and cloned in sorghum and maize, and a number of genes have been functionally studied. However, given the large genomes and complex regulatory mechanisms of maize and sorghum, it is difficult to elucidate the mechanism that controls photoperiod sensitivity response in C_4 crops.

Foxtail millet is an ideal model species for the study of photoperiod response in C_4 crops owing to its high inbreeding rate, C_4 photosynthesis, small diploid genome, and abundant

Abbreviations: qRT-PCR, quantitative real-time PCR; WT, wild type; GFP, green fluorescent protein; QTL, quantitative trait locus; Y2H, yeast two-hybrid; BiFC, bimolecular fluorescence complementation; HD, heading date; PL, panicle length; PW, panicle weight.

genetic resources with different sensitivities to photoperiod. Studies on photoperiod regulation in foxtail millet are scarce compared with those on maize and sorghum. To date, a small number of studies have been conducted on QTL mapping for photoperiod sensitivity-related traits in foxtail millet (Jia et al., 2013; Mauro-Herrera et al., 2013; Doust et al., 2017; Ni et al., 2017; Zhang et al., 2017). Xie (2012) showed that QTLs for photoperiod sensitivity-related traits in a short-day environment are located on chromosome 4 of foxtail millet, and QTLs for such traits in a long-day environment are located on chromosomes 3 and 9 of foxtail millet. Jia et al. (2013) identified 512 loci associated with 47 agronomic traits from a genome-wide association analysis of foxtail millet. However, the genetic basis of the photoperiod response of foxtail millet remains poorly understood.

Nevertheless, a small number of QTLs and genes associated with photoperiod sensitivity response have been reported in foxtail millet. Therefore, in this study, a high-density genetic map was constructed by whole-genome re-sequencing, the QTLs were mapped with short intervals, and a crucial candidate gene was cloned and identified. The findings provide valuable genetic information for innovation of the germplasm resources of foxtail millet and novel gene resources to accelerate the breeding of broadly adapted foxtail millet cultivars.

Materials and methods

Plant materials and field trials in different environments

A recombinant inbred line (RIL) mapping population consisting of 313 F₈ individuals was derived from a cross between the foxtail millet cultivars “Canggu 3” and “Longgu 3.” “Longgu 3” is a cultivar from northwestern China sensitive to photoperiod, whereas “Canggu 3” is from North China insensitive to photoperiod. The RIL population was generated using the single seed descent method.

Field trials for evaluation of the phenotypic performance of the RIL population were conducted in four environments (location/years or months) in China. The four environments comprised two locations: Jinan Experimental Station in Shandong Province and the Sanya Experimental Station in Hainan Province. The Jinan Experimental Station was adopted for field trials in May–September 2017, June–October 2017, and June–October 2018, and the Sanya Experimental Station was adopted for field trials in November–January 2017. These location/year or month combinations were designated 2017CQ, 2017XQ, 2018Q, and 2017HQ, respectively. A total of two replications were included in all environments using a randomized complete block design. Every line was grown in a two-row plot of length 500 cm, with a spacing of 3 cm between individuals and 50 cm between rows.

Measurement and analysis of phenotypic data for photoperiod sensitivity-related traits

For QTL analysis of photoperiod sensitivity-related traits, previous research has shown that heading date (HD), panicle length (PL), and panicle weight (PW) are suitable indicator traits for photoperiod sensitivity evaluation of foxtail millet (Jia et al., 2018). Therefore, HD, PL, and PW were recorded in each environment. In all environments, the phenotyping evaluation standard was identical. HD was defined as the period from the day of seedling emergence to the day when more than half of the individuals in a plot attained heading. PL was defined as the distance from the base to the apex of the main panicle. In total, five plants were randomly selected for evaluation of each phenotypic trait. The mean of the two replicates (blocks) for each phenotypic trait was used for QTL analysis.

Phenotype statistical analysis

The heritability, frequency distributions, and coefficient of variance (CV) were analyzed using SPSS Statistics version 17.0. The broad-sense heritability (H^2) of the photoperiod sensitivity-related traits was calculated in accordance with the method of Knapp et al. (1985) using the formula $H^2 = \delta_g^2 / (\delta_g^2 + \delta_{ge}^2/e + \delta^2/e \times r)$, where δ_g^2 is the genotype \times environment interaction, δ_{ge}^2 is the genetic variance, δ^2 is the error variance, r is the number of replications per environment, and e is the number of environments.

High-density genetic map construction

To construct DNA libraries for the RIL population, fresh leaves were sampled from each individual plant and immediately stored at -80°C until use. Genomic DNA was extracted using the Plant Genomic DNA Kit (TIANGEN, Beijing, China) in accordance with the manufacturer's protocol.

A re-sequencing strategy was adopted for high-density single-nucleotide polymorphism (SNP) development and RIL population genotyping. The genome of each RIL was re-sequenced on an Illumina HiSeq 2500 platform, and the raw data were analyzed by Illumina CASAVA 2.17 (Levy and Myers, 2016). Low-quality reads were filtered, and the remaining data were used for SNP calling. The clean reads were aligned to the reference genome of foxtail millet using the Burrows–Wheeler Aligner (Mckenna et al., 2010). The alignments were further analyzed and converted to BAM files for SNP calling and genotyping with SAMtools software (Li and Durbin, 2009). The SNPs were annotated using the Genome Analysis Toolkit (Mckenna et al., 2010), Picard, and SnpEff (Depristo et al., 2011; Cingolani et al., 2012).

A sliding window method was adopted to calculate the ratio of SNP alleles from “Longgu 3” and “Canggu 3” (Huang et al., 2009). The genotype data were scanned with 15 SNPs per window and a step size of 1 SNP. Consecutive SNPs with the same genotype were gathered into blocks. The locus between two different genotype blocks was determined to be the recombination breakpoint. A bin marker was defined as a consecutive 10-kb interval in which a recombination event was absent. A high-density genetic linkage map was constructed by mapping the bin markers on the nine foxtail millet chromosomes using HighMap (Liu et al., 2014).

Quantitative trait locus identification and candidate gene prediction

For mapping the QTLs associated with the photoperiod sensitivity-related traits, composition interval mapping and the “R/qtl” package were used. A logarithm of the odds (LOD) score greater than 2.5 was chosen as the threshold for declaration of the presence of a putative QTL in a particular region. The confidence interval was the region in which the LOD value intersected the threshold line.

The genes located in the confidence intervals were annotated with BLASTX (Altschul et al., 1997) by searching NR, COG, GO, and SwissProt 9 databases. Based on amino acid similarity, candidate genes located in all QTL intervals showing homology to genes related to photoperiod response in *Arabidopsis* or rice were identified, and putative *cis*-acting elements of the candidate genes were analyzed. Candidate genes located in the major QTL intervals with variations in the CDS region between two parents were identified.

Based on the RNA-seq data obtained,¹ the relative expression levels of all candidate genes were assessed in Yugu1 and A10. Yugu1 is insensitive to photoperiod, while A10 is sensitive to photoperiod. Furthermore, previous research suggested that the response of A10 to photoperiod was varying in different growth stages, that is, at the 1- to 3-leaf stage, the photoperiod response is the most sensitive, and then the sensitive effect decreased obviously at the 4- to 11-leaf stage, whereas when A10 reached the 12-leaf stage, the photoperiod response becomes the most insensitive. Therefore, we selected RNA-seq data from the samples on the first leaf at the seedling stage (the sensitive phase of photoperiod response) and on the flag leaf at the booting stage (the insensitive phase of photoperiod response) for expression pattern analysis.

Furthermore, several important candidate genes were selected for expression analysis by qRT-PCR in 20 millet varieties having dissimilar photoperiod sensitivities, and the

information of millet varieties is given in **Supplementary Table 5**. The most prospective candidate genes were shortlisted by the integration and overall consideration of the results from gene annotation, SNP and indel analysis, *cis*-element analysis, and expression analysis of candidate genes.

Cloning and function analysis of the candidate gene *SiCOL5*

The *SiCOL5* gene sequence was downloaded from the Phytozome database. The *SiCOL5* cDNA was amplified, and then the PCR products were purified and inserted into the pCambia1300 vector. The recombination vector was introduced into *Agrobacterium tumefaciens* strain GV3101 and then transformed into *Arabidopsis* by using the floral dip method. Transgenic plants were selected by culture on plates supplemented with hygromycin and then were transferred to soil. The transgenic plants used for phenotypic analysis were T₃ lines.

For expression characterization analysis of *SiCOL5*, fresh intact roots, stems, flag leaves, leaf sheaths, and panicles were collected. For analysis of circadian expression, fresh leaf samples were collected at 3-h intervals over a 48-h period from plants growing under short-days (light 9 h, dark 15 h) or long-days (light 15 h, dark 9 h). All samples were collected with three biological replicates. The sampled leaves were immediately frozen in liquid N₂ and stored at −80°C until use. The expression pattern of *SiCOL5* was analyzed by quantitative real-time PCR (qRT-PCR). The relative expression level of *SiCOL5* was calculated using the $2^{-\Delta\Delta Ct}$ method.

To analyze the function and regulatory network of *SiCOL5*, yeast two-hybrid (Y2H) assays were conducted. The fusion expression vectors pGADT7-*SiCOL5* and PGBKT7-SiNF-YA1 were constructed, and then the fusion expression vector pGADT7-*SiCOL5* + PGBKT7-SiNF-YA1, the positive vector pGADT7-T + PGBKT7-53, and negative control vector pGADT7-T + PGBKT7-lam were transformed separately into yeast (*Saccharomyces cerevisiae*) strain AH109 by using the polyethylene glycol/lithium acetate method. The yeast strains were cultured on the SD-T-L and SD-T-L-H-A plates for assessment of interaction between *SiCOL5* and SiNF-YA1 in yeast.

To analyze the subcellular location of the *SiCOL5* protein and further examine interaction between *SiCOL5* and SiNF-YA1, the fusion expression vectors p35S-*SiCOL5*-GFP, pSAT1-*SiCOL5*-nEYFP, and pSAT1-SiNF-YA1-cEYFP were generated by using a homologous recombination method. Protoplasts were isolated from rice leaf cells and were transformed with the fusion expression vector p35S-*SiCOL5*-GFP + p35S-OsGhd7-GFP, the control vector GFP + p35S-OsGhd7-GFP, or the pSAT1-*SiCOL5*-nEYFP + pSAT1-SiNF-YA1-cEYFP vector. The transformed protoplasts were incubated for 16–18 h in the dark.

¹ <http://foxtail-millet.biocloud.net/home>

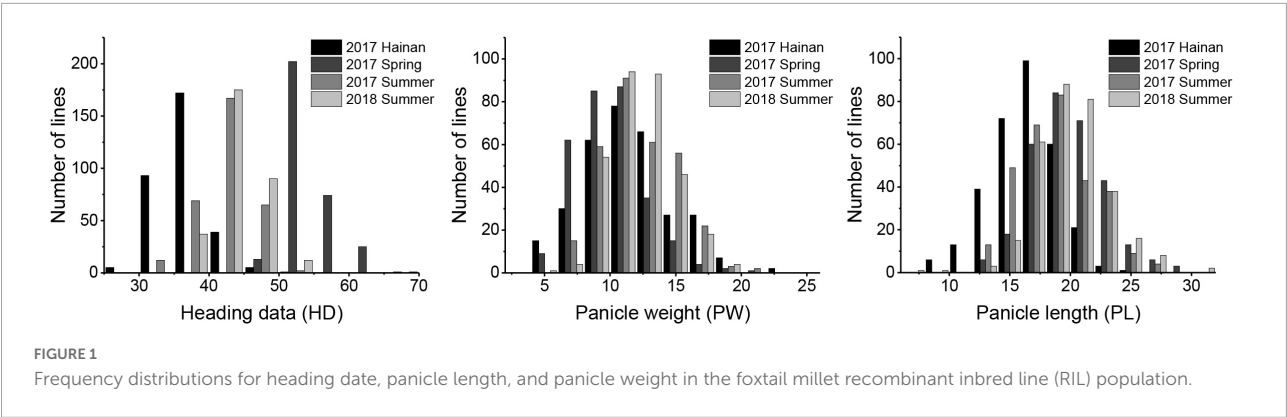


TABLE 1 Summary of phenotype data for quantitative traits of the foxtail millet recombinant inbred line (RIL) population in four environments.

Trait	Environments	Parents		RILs							
		M	P	Max.	Min.	Mean	SD	Skewness	Kurtosis	CV (%)	H ² (%)
Panicle length (cm)	17CQ	19.77	21.25	29.50	12.30	19.73	2.94	0.35	0.24	14.12	88.99
	17XQ	16.75	20.45	26.78	11.10	18.71	3.00	0.33	−0.36	16.03	
	17HQ	17.05	9.64	24.30	8.30	16.35	2.64	−0.30	0.31	16.14	
	18XQ	22.60	25.67	31.75	7.58	19.93	2.92	0.10	1.97	14.66	
Panicle weight per plant (g)	17CQ	12.18	6.52	21.37	4.45	10.08	2.66	0.78	1.30	26.38	87.48
	17XQ	12.40	9.63	21.31	6.25	12.03	2.72	0.37	−0.15	22.62	
	17HQ	12.85	5.14	23.83	4.00	11.47	3.39	0.29	0.46	29.54	
	18XQ	13.80	10.36	19.62	4.79	12.21	2.38	0.32	0.17	19.49	
Heading date (days)	17CQ	53	50	68	47	53.45	3.59	1.15	0.97	6.72	93.06
	17XQ	42	42	51	33	41.56	3.55	−0.11	−0.24	8.54	
	17HQ	39	26	53	26	36.51	3.68	0.51	1.14	10.09	
	18XQ	43	41	69	36	43.13	3.57	1.45	8.01	8.29	

M: “Canggu 3”; P: “Longgu 3.” H²: broad-sense heritability. CV, coefficient of variation.

Signals from the green fluorescent protein (GFP) and the yellow fluorescent protein (YFP) were observed using a confocal laser scanning microscope (FLV1200, Olympus, Tokyo, Japan).

These results indicated that the three traits were largely under heritable genetic control.

Results

Phenotype identification and analysis

“Longgu 3” and “Canggu 3” differed significantly in photoperiod sensitivity-related traits. Considerable phenotypic variation was observed in HD, PL, and PW in the RIL population across the four environments. Phenotypic values of HD, PW, and PL showed a normal distribution curve, which indicated that HD, PW, and PL are governed by multiple genes (Figure 1). Statistical analysis of the data for each phenotypic trait is presented in Table 1 and Supplementary Table 1. The CV for the three traits ranged from 6.72 to 29.54%. The broad-sense heritability of the three traits ranged from 87.48 to 93.06%.

Basic characteristics of the genetic map

A total of 2,076 bins were determined, which suggested that many recombination events had occurred in the RIL population (Supplementary Figure 1). A high-density genetic linkage map was constructed by mapping the 2,076 bin markers onto the nine foxtail millet chromosomes. The total genetic distance of the constructed linkage map was 1,417.97 cM, and the average distance between adjacent markers was 0.69 cM. The map contained 107,396 SNP loci and 2,076 bin markers. The average distance between bin markers ranged from 0.43 to 1.33 cM on a single chromosome (Table 2). Collinearity analysis showed that the linkage map was highly accurate, and Spearman correlation coefficients were greater than 0.99.

TABLE 2 Description of the basic characteristics of the genetic map for foxtail millet.

Linkage group ID ^a	Bin marker	Total distance (cM) ^b	Average distance (cM)
Chr1	325	179.90	0.56
Chr2	293	129.91	0.44
Chr3	155	140.38	0.91
Chr4	214	172.11	0.81
Chr5	200	183.38	0.92
Chr6	142	188.08	1.33
Chr7	418	179.83	0.43
Chr8	142	127.85	0.91
Chr9	187	116.53	0.63
Total	2076	1417.97	0.69

^aChr: linkage group ID on each chromosome. ^bTotal distance is the total genetic distance of a linkage group.

Quantitative trait loci identification and prediction of candidate genes

A total of 21 QTLs for HD, PW, and PL were identified (Figure 2). The phenotypic variation explained (PVE) value ranged from 1.64% (qPW5) to 69.92% (qHD1) within a trait (Table 3). Further analysis showed that 12 QTLs were detected in more than two environments, which were viewed as stable QTLs (shown in red or blue in Supplementary Figure 2). The retained QTLs that were detected in a single environment were considered to be unique QTLs (shown in green in Supplementary Figure 2).

Among the QTLs, five robust and major ones (PVE > 10%) were selected, but qHD2-2 and qPL2 were at the same locus (Supplementary Table 2). The QTL qHD1 had the largest effect and had 69.92 and 43.01% of the phenotypic variation

in 2017CQ and 2017XQ, respectively. This region contained 20 annotated genes. The QTL qHD2-2 had a strong influence on HD and had more than 10% of the phenotypic variation in all four environments. It is noteworthy that qHD2-2 was not only significant but also pleiotropic, and this region was associated with HD, PL, and PW. In a QTL analysis of the three traits combined, a crucial region was located adjacent to the Block14205–14218, corresponding to a physical distance of approximately 0.01 Mb (2017CQ and 2017XQ) to 0.07 Mb (2018Q) (Table 3), and this region contained 21 genes. The detailed list of the candidate genes of qHD1 and qHD2-2 can be found in Supplementary Table 3.

In all QTL intervals, a total of 2,803 genes were annotated by using the GO, COG, NR, and SwissProt databases. Overall, 19 photoperiod sensitivity-related candidate genes were predicted based on the gene annotations (Supplementary Table 4). The candidate genes were assigned to nine gene families, including the CCT family, NF-Y transcription factor family, and MADS-box family (Supplementary Figure 3). Analysis of putative *cis*-elements in the candidate genes indicated that most of the candidate genes had light-responsive elements (Supplementary Figure 4).

In the major QTL interval, a total of 270 genes were annotated, among which 97 genes had variations in the CDS region (Supplementary Table 2). In the qHD1 interval, Seita.1G021700 and Seita.1G021200, coding for uncharacterized protein, had non-synonymous mutation. In the qHD2-2 interval, Seita.2G445400 and Seita.2G444200 had frame shift variation, the function of which has not been characterized in millet. The variant of Seita.2G444300 between parents can be found in many regions, including CDS, upstream, intron, and downstream. Seita.2G444300 is a two-component response regulator-like PRR37 gene according to the Nr annotation, so we

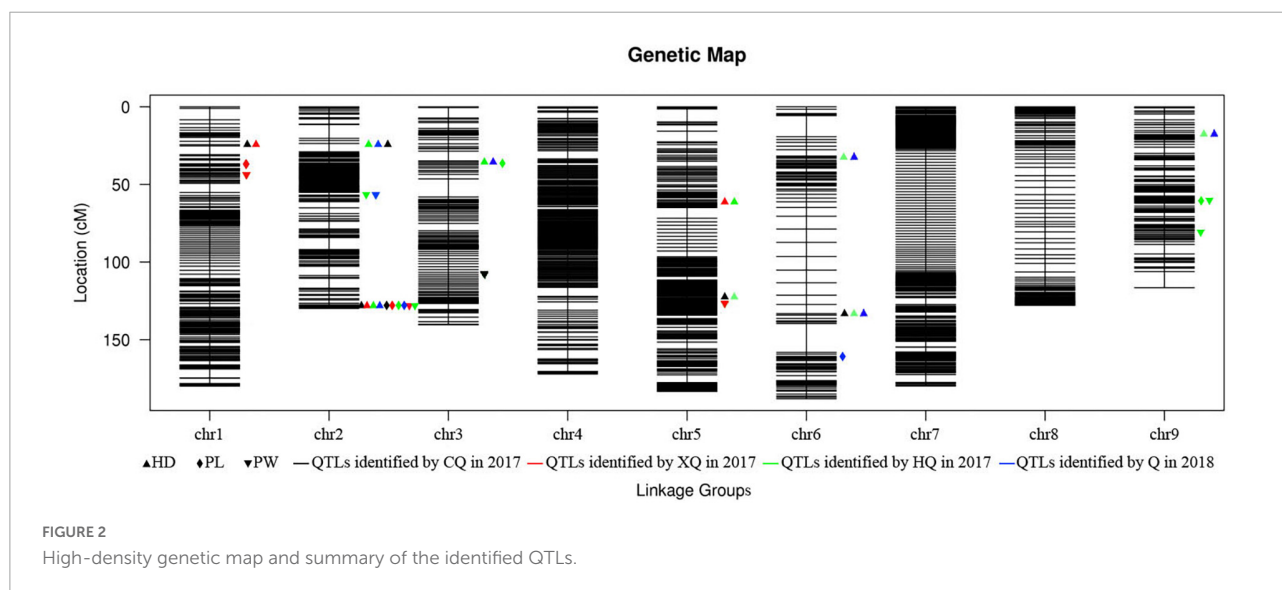


TABLE 3 Quantitative trait locus (QTL) mapping of three quality traits in the four environments.

QTL	Chr ^a	Marker	Genetic position (cM)			Physical position (Mb)			LOD ^b	PVE (%) ^c	ADD ^d	Environment
			Start	End	Distance	Start	End	Distance				
qHD1	1	Block544-545	18.93	18.93	0.00	1821152	2021933	0.20	5.18	69.92	−3.97	17CQ
		Block544-545	18.93	18.93	0.00	1821152	2021933	0.20	2.79	43.01	−3.08	17XQ
qHD2-1	2	Block5837-5914	20.32	23.62	3.29	2849742	3018544	0.17	3.09	0.75	0.41	17CQ
		Block5762-8152	11.14	30.82	19.68	2015203	8716474	6.70	4.86	6.32	1.21	17HQ
		Block5914	23.62	23.62	0.00	2994608	3018444	0.02	5.74	1.32	0.54	18Q
qHD2-2	2	Block14217	129.27	129.27	0.00	49169304	49182261	0.01	23.79	15.01	−1.84	17CQ
		Block14217	129.27	129.27	0.00	49169304	49182261	0.01	24.38	14.48	−1.79	17XQ
		Block14218	129.91	129.91	0.00	49183322	49200481	0.02	15.60	13.40	1.76	17HQ
		Block14205-14214	128.44	128.73	0.29	49095683	49166432	0.07	26.76	15.09	−1.84	18Q
qHD3	3	Block17835-17407	34.90	35.23	0.32	14213058	15842258	1.63	6.40	24.61	−2.39	17HQ
		Block17835-17407	34.90	35.23	0.32	14213058	15842258	1.63	4.03	10.19	−1.51	18Q
qHD5-1	5	Block29236-29591	50.20	64.94	14.75	7699919	9843267	2.14	5.68	2.43	0.73	17XQ
		Block29236-29591	50.20	64.94	14.75	7699919	9843267	2.14	5.08	5.13	1.09	17HQ
qHD5-2	5	Block29853-30143	111.56	125.13	13.58	13960422	24201025	10.24	2.67	0.63	0.38	17CQ
		Block29853-30211	111.56	134.24	22.69	13960422	24939701	10.98	3.25	4.96	1.07	17HQ
qHD6-1	6	Block33509	31.38	31.38	0.00	2042285	2052917	0.01	4.78	5.32	1.11	17HQ
		Block33116-34243	25.39	39.52	14.13	1388998	3280815	1.89	3.71	0.61	0.37	18Q
qHD6-2	6	Block36389-36421	137.85	138.34	0.48	27488651	28190802	0.70	4.50	5.50	−1.11	17CQ
		Block36389-36454	137.85	138.66	0.80	27488651	28255863	0.77	4.21	17.86	−2.04	17HQ
		Block36389-36392	137.85	138.01	0.16	27488651	27561418	0.07	2.83	3.09	−0.83	18Q
qHD9	9	Block60743-60773	20.06	20.28	0.22	5875509	5986760	0.11	2.66	3.97	0.96	17HQ
		Block60623-61055	13.33	24.19	10.87	4644491	8508650	3.86	3.54	1.48	0.56	18Q
qPL1	1	Block640-647	36.70	37.19	0.48	3496523	3770570	0.27	2.97	5.62	−0.94	17XQ
qPL2	2	Block14199-14214	128.02	128.73	0.70	48989024	49166432	0.18	9.98	8.18	−1.12	17CQ
		Block14217	129.27	129.27	0.00	49169304	49182261	0.01	26.00	19.94	−1.77	17XQ
		Block14195-14205	127.64	128.44	0.80	48946140	49116719	0.17	8.06	7.36	0.94	17HQ
		Block14205-14214	128.44	128.73	0.29	49095683	49166432	0.07	15.40	12.02	−1.33	18Q
qPL3	3	Block17614-17938	37.34	37.34	0.00	14964445	17317622	2.35	3.49	32.99	−1.99	17HQ
qPL6	6	Block37109-37120	176.34	177.04	0.71	32605531	32710929	0.11	4.43	4.00	−0.77	18Q
qPL9	9	Block61962-62010	57.49	57.97	0.48	40159357	40514420	0.36	3.44	3.97	0.69	17HQ
qPW1	1	Block890-993	48.95	49.27	0.32	5552939	6153865	0.60	2.71	2.75	−0.60	17XQ
qPW2-1	2	Block12177-12193	56.87	57.09	0.22	28277872	28383815	0.11	2.84	3.02	0.78	17HQ
		Block12177-12196	56.87	57.32	0.45	28277872	28420257	0.14	3.80	2.60	0.51	18Q
qPW2-2	2	Block14217-14218	129.27	129.91	0.64	49169304	49200481	0.03	6.20	4.50	−0.77	17XQ
		Block14217-14218	129.27	129.91	0.64	49169304	49200481	0.03	6.24	6.67	1.15	17HQ
qPW3	3	Block19556-19559	122.62	123.07	0.45	41416024	41855285	0.44	3.06	2.68	0.58	17CQ
qPW5	5	Block30267-30250	136.36	136.81	0.45	25373723	25484206	0.11	4.27	1.64	0.46	17XQ
qPW9-1	9	Block61951-61985	57.49	57.49	0.00	40159357	40321677	0.16	2.87	3.48	0.83	17HQ
qPW9-2	9	Block63401-63509	79.05	80.99	1.94	49748436	51702279	1.95	3.01	3.31	0.81	17HQ

^aChr: chromosome. ^bLOD, logarithm of the odds. ^cPVE, phenotypic variance explained. ^dADD, additive effect.

regarded Seita.2G444300 as the most probable candidate gene in the qHD2-2 interval. In the qHD3 interval, Seita.3G205400 and Seita.3G195900 contained premature stop codons; nevertheless, they code for uncharacterized protein. There were eight genes with frame shift variation. In the qHD6-2 interval, Seita.6G155600 and Seita.6G15860 contained a premature stop codon, and Seita.6G155600 is a cytochrome P450 71A1-like protein. Seita.6G155500 and Seita.6G155500 have frame shift variation. The detailed variation information between parents is given in [Table 4](#).

The expression patterns of all candidate genes were analyzed using RNA-seq data from Yugu1 and A10. The results showed that Seita.1G065300, Seita.2G444300, Seita.5G178100, Seita.9G468100, Seita.3G195200,

Seita.6G156100, Seita.3G1956100, and Seita.3G198500 had different expression patterns at the seedling stage between Yugu1 and A10. The expression levels of Seita.1G065300, Seita.2G444300, Seita.5G178100, Seita.9G468100, Seita.3G195200, and Seita.6G156100 were higher in A10 than those in Yugu1; on the contrary, the expression levels of Seita.3G195600 and Seita.3G198500 were lower in A10 than those in Yugu1. At the booting stage, the expression differences of Seita.1G065300, Seita.2G444300, Seita.3G198500, and Seita.6G156100 between Yugu1 and A10 were attenuated or insignificant; the expression profiles of these genes agreed with the result that the 12-leaf stage is insensitive to photoperiod response ([Supplementary Figure 5](#)).

TABLE 4 Variation analysis between parents of the candidate genes in major QTLs.

QTL	Gene ID	Nr annotation	Position	Ref	P	M	Variation type
qHD1	Seita.1G021700	Uncharacterized protein	1904650	G	A	C	Non-synonymous coding
	Seita.1G021200	Uncharacterized protein	1831278	C	C	C	Non-synonymous coding
qHD2-2	Seita.2G445400	Uncharacterized protein	49185509	TAG	T,TAG	TAG	Frame shift
	Seita.2G444200	Uncharacterized protein	49117637	AT	A	AT	Frame shift
	Seita.2G443600	Glycosyltransferase	49098978	C	CGCTGCT	C	Codon insertion
	Seita.2G443700	Uncharacterized protein	49100803	GAAC	GAAC	G	Codon insertion
	Seita.2G445100	Uncharacterized protein	49172292	G	C	G	Non-synonymous coding
	Seita.2G445200	Uncharacterized protein	49175179	C	T	C	Non-synonymous coding
	Seita.2G444700	Hypothetical protein SETIT_031465mg	49148770	G	A	G	Non-synonymous coding
	Seita.2G444900	Telomeric repeat-binding factor 2	49160124	C	A	C	Non-synonymous coding
	Seita.2G444600	Protein NETWORKED 2D	49142634	T	C	T	Non-synonymous coding
	Seita.2G444400	Probable serine/threonine-protein kinase PBL7	49138518	C	A	C	Non-synonymous coding
	Seita.2G443700	Uncharacterized protein LOC101756159	49100311	A	G	A	Non-synonymous coding
	Seita.2G443500	Phosphoinositide phospholipase C 2 isoform X2	49095673	G	C	G	Non-synonymous coding
	Seita.2G444800	Hypothetical protein SETIT_033421mg, partial	49150047	A	T	A	Non-synonymous coding
	Seita.2G444300	Like PRR37	49129307	G	A	G	Non-synonymous coding
qHD3	Seita.3G205400	Hypothetical protein SETIT_004917mg, partial	15836737	G	A	G	Stop gain
	Seita.3G195900	Uncharacterized protein	14896536	G	A	G	Stop gain
	Seita.3G189400	22.3 kDa class VI heat shock protein	14352409	CT	C,CT	CT	Frame shift
	Seita.3G197900	Hypothetical protein PAHAL_C02536	15000739	GT	G	GT	Frame shift
	Seita.3G195100	Uncharacterized protein	14839859	CCA	C	CCA	Frame shift
	Seita.3G199600	Uncharacterized protein	15101694	T	TA	T	Frame shift
	Seita.3G187700	H/ACA ribonucle protein complex NAF1	14228108	G	GCC,G	G	Frame shift
	Seita.3G198500	PREDICTED: ADP-ribosylation factor 2 isoform X1	15030744	CCTCT	C	CCTCT	Frame shift
	Seita.3G198600	Lipid transfer protein GPI-anchored 2-like isoform	15038941	C	CT	C	Frame shift
	Seita.3G195600	Myb family transcription factor PHL7	14870485	T	TA	T	Frame shift
	Seita.3G188400	Universal stress protein PHOS32	14266230	C	C,C	C	Codon deletion
	Seita.3G197000	Uncharacterized protein	14948561	GAGA	G	GAGA	Codon deletion
	Seita.3G203900	Probable WRKY transcription factor 63	15637813	T	TAGAGCATCC	T	Codon insertion
	Seita.3G204200	LRR receptor-like serine/threonine-protein kinase	15699729	G	GGGA,G	G	Codon insertion
	Seita.3G195300	Transcription factor TGAL1	14850912	A	G	A	Non-synonymous coding
	Seita.3G190100	Hypothetical protein SETIT_025028mg	14374820	G	A	G	Non-synonymous coding
	Seita.3G198900	NADH-cytochrome b5 reductase 1	15057250	T	C	T	Non-synonymous coding
	Seita.3G195000	Disease resistance protein RGA2 OS	14841648	G	A	G	Non-synonymous coding
qHD6-2	Seita.6G155600	Cytochrome P450 71A1-like	27559673	G	G	A	Stop gain
	Seita.6G158600	Uncharacterized protein	28097208	T	T	A	Stop gain
	Seita.6G155500	Uncharacterized protein	27556232	G	G	GCA	Frame shift
	Seita.6G159200	Uncharacterized protein	28214989	T	T	TC	Frame shift
	Seita.6G158900	Myb family transcription factor PHL8	28185819	T	T	TGACGCG	Codon insertion
	Seita.6G155500	Uncharacterized protein	27553999	A	A	G	Non-synonymous coding
	Seita.6G158900	Uncharacterized protein	28185596	C	C	T	Non-synonymous coding
	Seita.6G157000	Uncharacterized protein	27794643	G	A	G	Non-synonymous coding
	Seita.6G156500	Uncharacterized protein	27695570	A	A	G	Non-synonymous coding
	Seita.6G159400	Uncharacterized protein	28223049	T	T	C	Non-synonymous coding

Furthermore, six candidate genes, namely, Seita.1G065300, Seita.2G444300, Seita.9G468100, Seita.5G178100, Seita.2G445400, and Seita.3G195900, were selected for expression analysis by qRT-PCR in 20 millet varieties which had dissimilar photoperiod sensitivities. The qRT-PCR results showed that Seita.1G065300, Seita.2G444300, and Seita.9G468100 had different expression patterns in two categories of varieties having different photoperiod sensitivities. The expression patterns of Seita.1G065300, Seita.2G444300, and Seita.9G468100 were similar, and they all had higher expression levels in millet varieties sensitive to photoperiod. This result agreed with the RNA-seq data. It is interesting to

note that the expression of Seita.2G444300 was not detected in all selected millet varieties insensitive to photoperiod but only in four varieties sensitive to photoperiod. The expression level of Seita.2G445400 and Seita.5G178100 was low in all selected millet varieties. The expression of Seita.3G195900 is irregular in the two categories of millet varieties which had different photoperiod sensitivities (Figure 3).

In summary, among the candidate genes, Seita.1G065300, Seita.2G444300, and Seita.9G468100 were selected as important candidate genes for photoperiod response by the integration and overall consideration of the results from gene annotation, SNP and indel analysis, *cis*-element analysis, and the expression

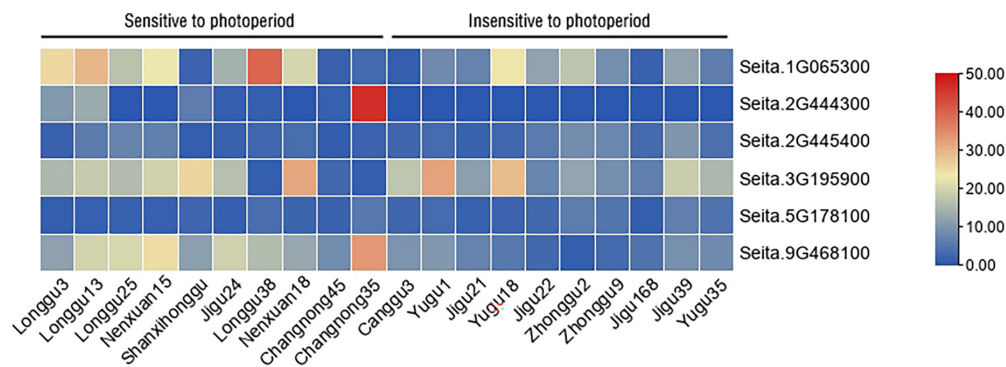


FIGURE 3

Expression analysis of important candidate genes in different millet varieties having different photoperiod sensitivities. Different colors in the heatmap represent gene transcript values.

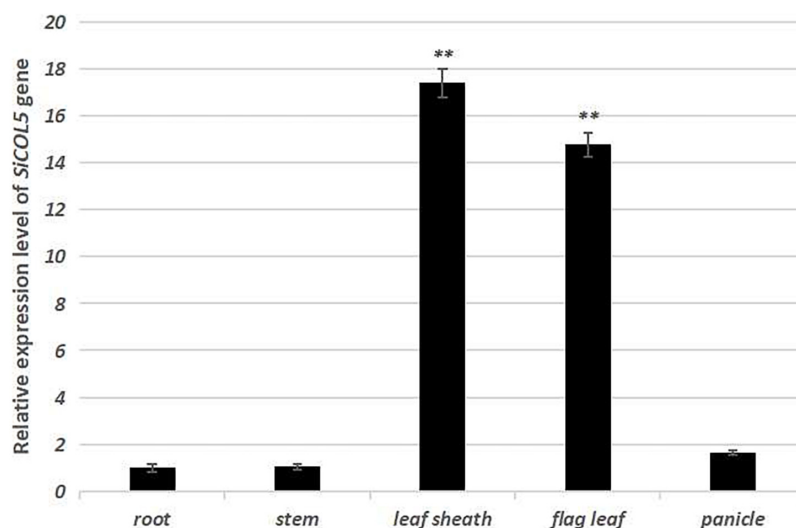


FIGURE 4

Expression patterns of the candidate gene *SiCOL5* in various tissues of foxtail millet. ** $p < 0.01$.

pattern of different genes in different varieties having different photoperiod sensitivities.

Cloning and preliminary function analysis of the candidate gene *SiCOL5*

One important candidate gene was cloned and was designated *SiCOL5*. The gene encoded a 386-amino acid protein and belonged to the CCT family. The qRT-PCR analysis indicated that *SiCOL5* transcripts were more abundant in the flag leaf and leaf sheath than in other tissues (Figure 4), which was consistent with the public expression data. It also suggested that *SiCOL5* may perform a crucial function in the leaf. Furthermore, the expression of *SiCOL5* showed a circadian pattern in short-day and long-day environments

(Figure 5). However, the circadian expression of *SiCOL5* differed under the two photoperiods. These results suggested that *SiCOL5* was sensitive to photoperiod and was regulated by biological rhythms. The *SiCOL5* gene was ectopically expressed in *Arabidopsis* to investigate the function of *SiCOL5*. The transgenic *Arabidopsis* lines overexpressing *SiCOL5* produced fewer leaves and bloomed earlier (Figure 6) than wild-type *Arabidopsis*. The results suggested that *SiCOL5* positively regulated the flowering time in *Arabidopsis*.

Subcellular localization analysis showed that *SiCOL5* protein was localized in the nucleus (Figure 7A). A Y2H assay showed that *SiCOL5* was capable of interacting with SiNF-YA1 in yeast (Figure 7B). This was confirmed by performing a bimolecular fluorescence complementation (BiFC) assay in rice protoplasts. These results indicated that *SiCOL5* interacted with SiNF-YA1 in the nucleus (Figure 7C).

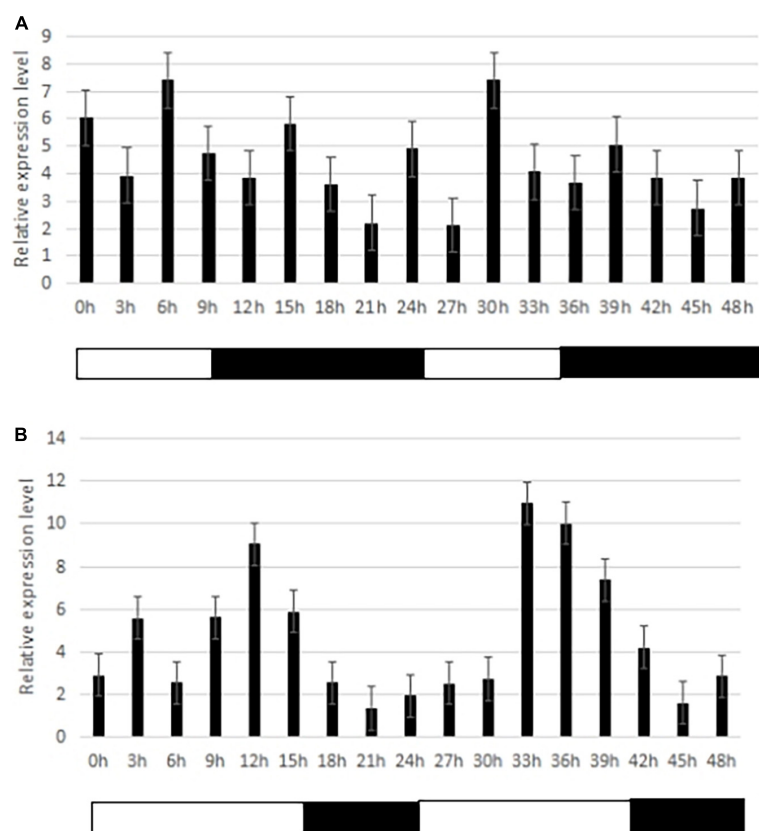


FIGURE 5

Circadian expression of the candidate gene *SiCOL5* under different photoperiod treatments. (A) Short-days; (B) long-days. Black bars represent the dark period, and white bars represent the light period.

Discussion

Cultivars that show broad adaptability in the commercial production and breeding of foxtail millet are scarce, which strongly limits further development and strengthening of the competitive power of the millet seed industry. Therefore, it is important to expand and strengthen the millet seed sector by improving the adaptability of foxtail millet cultivars (Jia and Diao, 2022). However, the genetic basis of the photoperiod sensitivity response of foxtail millet and the main effect genes involved in regulating the photoperiod sensitivity response remain elusive. Mapping and identification of the major genes that regulate the photoperiod sensitivity response are vital to achieve rapid and directional genetic improvement of cultivar adaptability.

The resolution of genetic maps has an important impact on the accuracy of QTL mapping. Compared with previously reported genetic maps of foxtail millet (Zhang et al., 2012; Qie et al., 2014), such as a random fragment length polymorphism-based map (Wang et al., 1998), a simple sequence repeat linkage map (Jia et al., 2009), and a genetic map constructed with 992 SNPs (Bennetzen et al., 2012), a greater number of markers

were used for construction of the present map, and the marker density and resolution of the present genetic map were higher. Therefore, a greater number of recombination events could be detected. Based on the high-density genetic map, narrowing of the regions between markers and prediction of candidate genes could be achieved more accurately.

In the present study, 21 QTLs were detected, but a majority of those were novel, and only a small number of those were consistent with previously detected QTLs. This may be due to differences in the environments or populations used between studies. Xie (2012) detected QTLs for photoperiod sensitivity-related traits on chromosome 4 of foxtail millet grown under short-days, and on chromosomes 3 and 9 of foxtail millet grown under long-days. However, we failed to detect corresponding QTLs on chromosome 4. The reason may be that Xie used flowering date as the phenotypic trait, whereas we selected HD as a photoperiod sensitivity-related trait. However, only highly heritable and stable QTLs can be used for marker-assisted selection. In this study, qHD1 and qHD2-2 were stable QTLs and had more than 10% of phenotypic variation. It is notable that Jia et al. (2013) also detected genomic regions showing strong association signals for HD on chromosomes 1

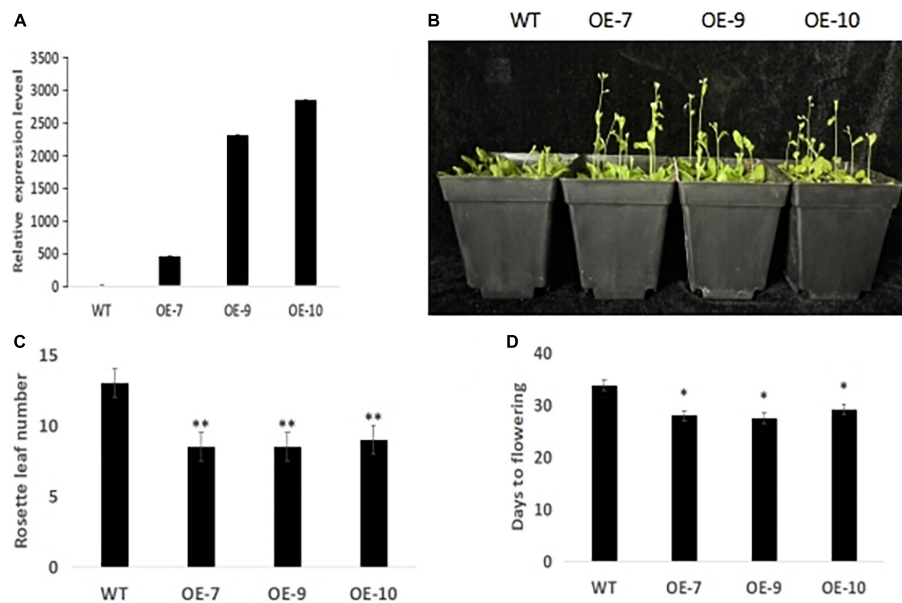


FIGURE 6

SiCOL5 positively regulates flowering time in *Arabidopsis*. (A) Identification of *SiCOL5* overexpression plants by qRT-PCR. (B–D) Phenotype (B), rosette leaf number (C), and days to flowering (D) of *SiCOL5* overexpression plants compared with wild-type plants. ** $p < 0.01$, * $p < 0.05$.

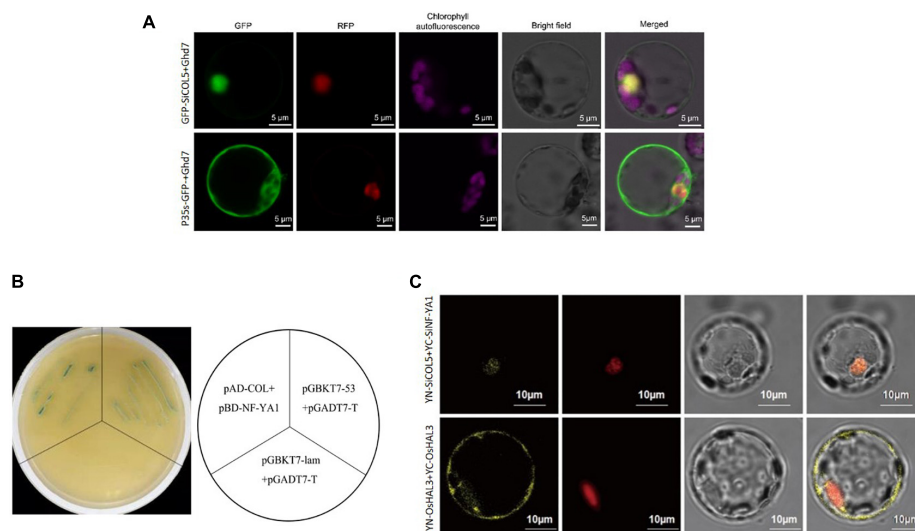


FIGURE 7

Interaction between *SiCOL5* and *SiNF-YA1* in the nucleus. (A) Subcellular localization of *SiCOL5*. (B,C) Interaction between *SiCOL5* and *SiNF-YA1* in a yeast two-hybrid assay (B) and bimolecular fluorescence complementation assay (C). Scale bar = 10 μ m.

and 2. Therefore, we suggest that qHD1 and qHD2-2 could be considered priority candidates for marker-assisted selective breeding.

In the present study, 116 candidate genes were predicted. Among the candidate genes, *Seita.2G085600* and *Seita.2G444300* also were mapped by genome-wide association by Jia et al. (2013). *Seita.2G444300* (*SiPRR37*) was

reported to be involved in photoperiod sensitivity in millet; therefore, the results verified the QTL mapping results of this study. In addition, most candidate genes identified in this study were novel genes, such as the candidate genes in the major QTL intervals. The QTL qHD1 had the largest effect and had 69.92 and 43.01% of phenotypic variation in 2017CQ and 2017XQ, respectively. In the qHD1 interval, *Seita.1G021700*

and Seita.1G021200 had non-synonymous mutation, and they were uncharacterized protein. In qHD2-2, Seita.2G445400 and Seita.2G444200 had frame shift variation, the function of which also has remained uncharacterized by now in millet. Therefore, the uncharacterized genes in the major QTL intervals that had important variants between parents should be prospective, which could be used for foxtail millet breeding.

We cloned *SiCOL5* following expression analysis of the candidate genes (Figure 3 and Supplementary Figure 5). Xilei et al. (2020) analyzed advances in research on molecular mechanisms of the photoperiod regulation of plant flowering and the *CCT* gene family. They suggested that the *CCT* family plays a critical role in photoperiod regulation of the flowering pathway. LinLin et al. (2021) reported that a significant genome-wide association signal (position: 31456761 bp) was detected on chromosome 1, and in this region, only one *CCT* family gene (*SiTOC1*) was identified. Further analysis identified many haplotypic variations of *SiTOC1*, but the REC and CCT domains of *SiTOC1* were conserved in foxtail millet accessions (LinLin et al., 2021). These results also suggested that the *CCT* domain plays a key role in photoperiod regulation of the flowering pathway.

In this study, some QTLs were pleiotropic or showed tight linkage with other QTLs, such as qHD2-2. This finding suggested that the regions with the pleiotropic QTLs may be hot spots that control photoperiod sensitivity response-related traits. Functional identification of the genes in the hot spot regions is important for breeding. In addition, many separate QTLs were mapped that controlled HD, PL, or PW, rather than two or three traits, which suggests that each trait should have a relatively independent genetic pathway. Therefore, fine mapping of the QTL is needed for QTL cloning. The present results provide an important basis for further fine mapping of QTLs and contribute to elucidation of the molecular mechanisms of photoperiod responses in foxtail millet.

In the current study, 12 QTLs for HD, PL, and PW were stable QTLs that were detected in multiple environments; thus, the QTLs were insensitive to different environments. There may be two reasons for this: the QTLs may have important effects on phenotypes, or the QTLs may have been overlooked via current statistical methods, although they exert a large genotype–environment interaction effect.

Conclusion

A high-density linkage map was constructed for foxtail millet based on large-scale development of SNPs and bin markers by re-sequencing. Using this high-density map, three photoperiod sensitivity-related phenotypic traits were mapped, and 21 QTLs were identified in four environments. In total, 116 candidate genes were predicted, and three major genes were selected. An important candidate gene, *SiCOL5*, was cloned and identified. These results provide valuable genetic information

for the innovation of germplasm resources of foxtail millet and novel gene resources to accelerate the breeding of foxtail millet cultivars with broad adaptability.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA853459.

Author contributions

F-FL and Y-AG coordinated the project, and conceived and designed the experiments. J-HN, XY, and Q-HK conducted the experiments. R-FW conducted the bioinformatics work. F-FL wrote the first draft. LQ, E-YC, and Y-BY edited the manuscript. L-NL, Z-YL, H-WZ, and H-LW contributed valuable discussions. All authors have read and approved the final manuscript version.

Funding

This study was financially supported by the Agricultural Fine Seed Project of Shandong Province (2021LZGC006), the China's Agricultural Research System (CARS-06-14.5-A19), the Open Project of the National Engineering Research Center of Wheat and Maize, and the Key Research and Development Project of Shandong Province (2021LZGC025).

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.969604/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Recombination bin map for the foxtail millet recombinant inbred line (RIL) population. The bin map comprises 2076 bin markers. Physical position is based on Yugu1 RefGen-V2.2 sequence. Red: "Longgu 3" genotype; blue: "Canggu 3" genotype; yellow: heterozygote.

SUPPLEMENTARY FIGURE 2

Summary of QTL identifications. **(A)** QTL mapping for heading date (HD). **(B)** QTL mapping for panicle length (PL). **(C)** QTL mapping for panicle weight (PW). QTLs that were expressed stably in all four environments are shown in red. QTLs that were stable in two or three environments are shown in blue. QTLs that were specifically expressed in only one environment are shown in green.

SUPPLEMENTARY FIGURE 3

Classification of candidate genes associated with photoperiod sensitivity response in foxtail millet.

SUPPLEMENTARY FIGURE 4

Putative *cis*-element analysis of the candidate genes associated with flowering. Different *cis*-elements are indicated by different colors. Light-responsive elements, low temperature-responsive elements, and gibberellin-responsive elements were analyzed.

SUPPLEMENTARY FIGURE 5

Expression profiles of the candidate genes in Yugu1 and A10. Different colors in the heatmap represent gene transcript values.

References

- Altschul, S. F., Madden, T. L., and Schäffer, A. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389
- Anwer, M. U., Davis, A., Davis, S. J., and Quint, M. (2018). Photoperiod sensing of the circadian clock is controlled by ELF3 and GI. *Plant J.* 101, 1397–1410. doi: 10.1101/321794
- Bennetzen, J. L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A. C., et al. (2012). Reference genome sequence of the model plant *Setaria*. *Nat. Biotechnol.* 30, 555–561. doi: 10.1038/nbt.2196
- Chardon, F., Virlon, B., Moreau, L., Falque, M., Joets, J., Decousset, L., et al. (2004). Genetic architecture of flowering time in maize as inferred from quantitative trait loci meta-analysis and synteny conservation with the rice genome. *Genetics* 168, 2169–2185. doi: 10.1534/genetics.104.032375
- Cheng, H., Sun, H., Xu, D., Chen, Q., and Feng, T. (2018). ZmCCT9 enhances maize adaptation to higher latitudes. *Proc. Natl. Acad. Sci. U.S.A.* 115, E334–E341. doi: 10.1073/pnas.1718058115
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms SnpEff. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Coles, N. D., McMullen, M., Balint-Kurti, P. J., Pratt, R. C., and Holland, J. B. (2010). Genetic Control of Photoperiod Sensitivity in Maize Revealed by Joint Multiple Population Analysis. *Genetics* 184, 799–812. doi: 10.1534/genetics.109.110304
- Depristo, M. A., Banks, E., Poplin, R., Garimella, K. V., and Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498. doi: 10.1038/ng.806
- Doust, A. N., Margarita, M. H., Hodge, J. G., and Jessica, S. (2017). The C4 Model Grass *Setaria* Is a Short Day Plant with Secondary Long Day Genetic Regulation. *Front. Plant Sci.* 8, 1062–1063. doi: 10.3389/fpls.2017.01062
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* 19, 1068–1076. doi: 10.1101/gr.089516.108
- Jia, G., and Diao, X. (2022). Current Status and Perspectives of Innovation Studies Related to Foxtail Millet Seed Industry in China. *Sci. Agric. Sin.* 55, 653–665. doi: 10.3864/j.issn.0578-1752.2022.04.003
- Jia, G., Huang, X., Zhi, H., Zhao, Y., Zhao, Q., Li, W., et al. (2013). A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat. Genet.* 45, 957–961. doi: 10.1038/ng.2673
- Jia, X., Li, J., and Zhang, B. (2018). Comprehensive evaluation of main agronomic traits of Millet resources under different light and temperature conditions. *Sci. Agric. Sin.* 51, 2429–2441.
- Jia, X., Zhang, Z., Liu, Y., Zhang, C., Shi, Y., Song, Y., et al. (2009). Development and genetic mapping of SSR markers in foxtail millet [*Setaria italica* (L.) P. Beauv.]. *Theor. Appl. Genet.* 118, 821–829. doi: 10.1007/s00122-008-0942-9
- Jin, M., Liu, X., Wei, J., Liu, H., and Yan, J. (2018). ZmCOL3, a CCT gene represses flowering in maize by interfering circadian clock and activating expression of ZmCCT. *J. Integr. Plant Biol.* 60, 465–480. doi: 10.1111/jipb.12632
- Kazuyuki, Izawa, Takeshi, Fuse, Takuichi, Yamanouchi, et al. (2004). *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of *Hd1*. *Genes. Dev.* 18, 926–936. doi: 10.1101/gad.1189604
- Klein, R. R., Miller, F. R., Dugas, D. V., Brown, P. J., Burrell, A. M., and Klein, P. E. (2015). Allelic variants in the PRR37 gene and the human-mediated dispersal and diversification of sorghum. *Theor. Appl. Genet.* 128, 1669–1683. doi: 10.1007/s00122-015-2523-z
- Knapp, S. J., Stroup, W. W., and Ross, W. M. (1985). Exact Confidence Intervals for Heritability on a Progeny Mean Basis. *Crop Sci.* 25, 192–194. doi: 10.2135/cropsci1985.001183X002500010046x
- Lee, S., Kim, J., Han, J. J., Han, M. J., and An, G. (2004). Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF OVEREXPRESSION OF CO 1/AGAMOUS?IKE 20 (SOC1/AGL20) ortholog in rice. *Plant J.* 38, 754–764. doi: 10.1111/j.1365-3113X.2004.02082.x
- Levy, S. E., and Myers, R. M. (2016). Advancements in Next-Generation Sequencing. *Annu. Rev. Genom. Hum. Genet.* 17, 95–115. doi: 10.1146/annurev-genom-083115-022413
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- LinLin, Z., Sha, Z. H. T., RenLiang, Z., Wei, Z., GuanQing, J., and XianMin, D. (2021). Characterizations of Transcriptional and Haplotypic Variations of SiTOC1 in Foxtail Millet. *Sci. Agric. Sin.* 54, 2273–2286. doi: 10.3864/j.issn.0578-1752.2021.11.003
- Liu, D., Ma, C., Hong, W., Huang, L., Min, L., Liu, H., et al. (2014). Construction and Analysis of High-Density Linkage Map Using High-Throughput Sequencing Data. *PLoS One* 9:e98855. doi: 10.1371/journal.pone.0098855
- Mauro-Herrera, M., Wang, X., Barbier, H., Brutnell, T. P., Devos, K. M., and Doust, A. N. (2013). Genetic Control and Comparative Genomic Analysis of Flowering Time in *Setaria* (Poaceae). *G3* 3, 283–295. doi: 10.1534/g3.112.005207
- 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
- Miller, T. A., Muslin, E. H., and Dorweiler, J. E. (2008). A maize CONSTANS-like gene, conz1, exhibits distinct diurnal expression patterns in varied photoperiods. *Planta* 227, 1377–1388. doi: 10.1007/s00425-008-0709-1
- Murphy, R. L., Klein, R. R., and Morishige, J. (2011). Coincident light and clock regulation of pseudoresponse regulator protein 37 (PRR37) controls photoperiodic flowering in sorghum. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16469–16474. doi: 10.1073/pnas.1106212108
- Ni, X., Xia, Q., Zhang, H., Cheng, S., Li, H., Fan, G., et al. (2017). Updated foxtail millet genome assembly and gene mapping of nine key agronomic traits by resequencing a RIL population. *GigaScience* 6, 1–8. doi: 10.1093/gigascience/giw005
- Qie, L., Jia, G., Zhang, W., Schnable, J., Shang, Z., Li, W., et al. (2014). Mapping of Quantitative Trait Locus (QTLs) that Contribute to Germination and Early Seedling Drought Tolerance in the Interspecific Cross *Setaria italica* × *Setaria viridis*. *PLoS One* 9:e101868. doi: 10.1371/journal.pone.0101868
- Sim, S. A., Woo, S. G., Hwang, D. Y., Kim, J. H., and Song, Y. H. (2019). FLOWERING HTH1 is involved in CONSTANS-mediated flowering regulation in Arabidopsis. *Appl. Biol. Chem.* 62:56.
- Su, H., Cao, Y., Lixia, K., Wen, Y., Cao, Y., Ren, Z., et al. (2018). Dual functions of ZmNF-YA3 in photoperiod-dependent flowering and abiotic stress responses in maize. *J. Exp. Bot.* 69, 5177–5189. doi: 10.1093/jxb/ery299

- Tsuji, H., Tamaki, S., Komiya, R., and Shimamoto, K. (2008). Florigen and the Photoperiodic Control of Flowering in Rice. *Rice* 1, 25–35. doi: 10.1007/s12284-008-9005-8
- Wang, C. L., Cheng, F. F., Sun, Z. H., Tang, J. H., Wu, L. C., Ku, L. X., et al. (2008). Genetic analysis of photoperiod sensitivity in a tropical by temperate maize recombinant inbred population using molecular markers. *Theor. Appl. Genet.* 117, 1129–1139. doi: 10.1007/s00122-008-0851-y
- Wang, Z. M., Devos, K. M., Liu, C. J., Wang, R. Q., and Gale, M. (1998). Construction of RFLP-based maps of foxtail millet *Setaria italica* (L.) P. Beauv. *Theor. Appl. Genet.* 96, 31–36. doi: 10.1007/s001220050705
- Xie, L. (2012). *QTL Mapping and Analysis for the Related Traits of Photoperiod Sensitivity in Foxtail Millet*. Henan: Henan Agricultural University.
- Xilei, Y., Zhenshan, W., Xiaoping, J., Luman, S., Jianfeng, L., and Bo, Z. (2020). Research advances on molecular mechanisms of photoperiod-regulation plant flowering and CCT gene family. *Acta Agric. Zhejiangensis* 6, 1133–1140.
- Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., et al. (2009). Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat. Genet.* 40, 761–767. doi: 10.1038/ng.143
- Yang, S., Murphy, R. L., Morishige, D. T., Klein, P. E., Rooney, W. L., and Mullet, J. E. (2014). Sorghum Phytochrome B Inhibits Flowering in Long Days by Activating Expression of *SbPRR37* and *SbGHD7*. Repressors of *SbEHD1*, *SbCN8* and *SbCN12*. *PLoS One* 9:e105352. doi: 10.1371/journal.pone.0105352
- Yano, M. (2000). Hd1, a Major Photoperiod Sensitivity Quantitative Trait Locus in Rice, Is Closely Related to the Arabidopsis Flowering Time Gene CONSTANS. *Plant Cell* 12, 2473–2484. doi: 10.2307/3871242
- Zhang, G. Y. L., Xin, Z. W., and Quan, S. F. (2012). Genome sequence of foxtail millet (*Setaria italica*) provides insights into grass evolution and biofuel potential. *Nat. Biotechnol.* 30, 549–554. doi: 10.1038/nbt.2195
- Zhang, K., Fan, G., Zhang, X., Zhao, F., Wei, W., Du, G., et al. (2017). Identification of QTLs for 14 Agronomically Important Traits in *Setaria italica* Based on SNPs Generated from High-Throughput Sequencing. *G3* 7, 1587–1594. doi: 10.1534/g3.117.041517



OPEN ACCESS

EDITED BY

Rajeev K. Varshney,
Murdoch University, Australia

REVIEWED BY

Babu Motagi,
University of Agricultural Sciences,
India
Papa Rao Vaikuntapu,
Directorate of Groundnut Research
(ICAR-DGR), India
Han Xia,
Shandong Academy of Agricultural
Sciences, China

*CORRESPONDENCE

Guilan Li
lgl63@126.com
Zhangying Wang
wangzhangying@gdaas.cn

SPECIALTY SECTION

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

RECEIVED 14 July 2022

ACCEPTED 04 October 2022

PUBLISHED 20 October 2022

CITATION

Jia R, Zhang R, Gangurde SS, Tang C,
Jiang B, Li G and Wang Z (2022)
Integrated analysis of carotenoid
metabolites and transcriptome
identifies key genes controlling
carotenoid compositions and content
in sweetpotato tuberous roots
(*Ipomoea batatas* L.).
Front. Plant Sci. 13:993682.
doi: 10.3389/fpls.2022.993682

COPYRIGHT

© 2022 Jia, Zhang, Gangurde, Tang,
Jiang, Li 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) and the copyright
owner(s) 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.

Integrated analysis of carotenoid metabolites and transcriptome identifies key genes controlling carotenoid compositions and content in sweetpotato tuberous roots (*Ipomoea batatas* L.)

Ruixue Jia^{1,2}, Rong Zhang², Sunil S. Gangurde^{3,4,5},
Chaochen Tang², Bingzhi Jiang², Guilan Li^{1*}
and Zhangying Wang^{2*}

¹College of Agronomy and Biotechnology, Hebei Normal University of Science and Technology, Changli, China, ²Crops Research Institute, Guangdong Academy of Agricultural Sciences & Key Laboratory of Crop Genetic Improvement of Guangdong Province, Guangzhou, China, ³International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, ⁴Crop Protection and Management Research Unit, USDA-ARS, Tifton, GA, United States, ⁵Department of Plant Pathology, University of Georgia, Tifton, GA, United States

Sweetpotato (*Ipomoea batatas* L.) with different depths of yellow color contains different compositions of carotenoids, which are beneficial for human health. In this study, we performed an integrated analysis of metabolomic and transcriptomic to identify key genes playing a major role in carotenoid coloration in sweetpotato tuberous roots. Herein, 14 carotenoids were identified in five sweetpotatoes. Orange-red and orange cultivars were dominated by β -carotene (385.33 $\mu\text{g/g}$ and 85.07 $\mu\text{g/g}$), yellow cultivar had a high β -cryptoxanthin (11.23 $\mu\text{g/g}$), light-yellow cultivar was rich in zeaxanthin (5.12 $\mu\text{g/g}$), whereas lutein (3.34 $\mu\text{g/g}$) was the main carotenoid in white cultivar. Furthermore, 27 differentially expressed genes involved in carotenoid metabolism were identified based on comparative transcriptome. Weighted gene co-expression network analysis identified 15 transcription factors highly associated with carotenoid content in sweetpotatoes. These results provide valuable information for revealing the regulatory mechanism of carotenoid metabolism in different-colored sweetpotato tuberous roots.

KEYWORDS

sweetpotato, coloration, key genes, transcription factor, WGCNA

Introduction

Sweetpotato (*Ipomoea batatas* L.) ranks the sixth largest food crop in the world and has played an important role as a famine-relief crop in its long history (CIP, 2020). Recently, sweetpotato has been re-evaluated as a health-promoting food and successfully demonstrated to tackle vitamin A deficiency with biofortification in sub-Saharan Africa due to its high accumulation of β -carotene (World Food Prize Foundation, 2016). In Asia, yellow-fleshed cultivars are popular for table use and processed products. The flesh color of the sweetpotato has become one of the most important targets in breeding programs. Sweetpotato cultivars show various flesh colors such as white, yellow, orange and purple due to the accumulation of different contents of pigments. Among them, the yellow and orange flesh colors are determined by the carotenoid content and composition (Ishiguro et al., 2010). In addition to being an excellent source of provitamin A in humans, carotenoids also have the function of antioxidant activity, anticancer activity, reducing the progression of age-related macular eye disease in human health (Eggersdorfer and Wyss, 2018; Fraser and Bramley, 2004; Islam et al., 2016).

Due to the visibility and importance of carotenoid in plants and humans, its biosynthesis pathway and the associated biosynthetic enzymes have been thoroughly investigated (Fraser and Bramley, 2004; Sajilata et al., 2008). Geranylgeranyl diphosphate (GGPP) is a common precursor of carotenoids, ubiquinones, abscisic acid, tocopherols, gibberellin and chlorophyll (Beck et al., 2013). Carotenoid biosynthesis pathway consists of two processes to produce two different kinds of carotenoids. First, the GGPP is converted into a group of carotenoids, containing only carbon and hydrogen called as carotenes by a series of enzymes. Further, the carotenes are turn converted into other group of carotenoids, containing one or more oxygen atoms called as xanthophylls (Armstrong, 1997). Precursor GGPP produces colorless phytoene through

condensation catalyzed by 15-cis-phytoene synthase (PSY) and the phytoene produces red lycopene through a series of enzymes (PDS, 15-cis-phytoene desaturase; Z-ISO, 15-cis- ζ -carotene isomerase; ZDS, ζ -carotene desaturase; CRTISO, polycopene isomerase). Lycopene is the critical substance in the carotenoid biosynthesis pathway, which is then differentiates into two branches: γ -carotene and δ -carotene. In the δ -carotene branch, lutein is finally produced by lycopene ϵ -cyclase (LCYE), lycopene β -cyclase (LCYB), and carotenoid ϵ -hydroxylase (CYP97C1) and β -carotene hydroxylase (CHYB). In the γ -carotene branch, the carotenoids synthesized with the help of lycopene β -cyclase (LCYB), β -carotene hydroxylase (CHYB), zeaxanthin epoxidase (ZEP), violaxanthin de-epoxidase (VDE) (Liu et al., 2020; Suematsu et al., 2020; Xu et al., 2021; KEGG, 2022).

As mentioned above, cultivated sweetpotatoes have the ability to synthesize various kinds of carotenoids in tuberous root, which are responsible for the vivid flesh colors. The carotenoids in the tuberous root of sweetpotato include β -carotene, β -cryptoxanthin, zeaxanthin, violaxanthin and other carotenoids (Ishiguro et al., 2010; Kim et al., 2012; Kim et al., 2013; Kim et al., 2014). Ishiguro et al. (Ishiguro et al., 2010) analyzed the different colors of eight sweetpotato cultivars and breeding lines, and showed that the main carotenoids in five different yellow-fleshed sweetpotato are β -carotene 5,8;5',8'-diepoxide (32–51%) and β -cryptoxanthin 5,8-epoxide (11–30%), and β -carotene (< 10%). The main carotenoid in orange-fleshed sweetpotato is β -carotene (80–92%), whereas other carotenoids constitute less than 2% of the total in three different cultivars (Ishiguro et al., 2010). Some genes encoding enzymes in the carotenoid metabolic pathway, including PSY, PDS, ZDS, CRTISO, LCYB, LCYE, CHYB, ZEP, 9-cis-epoxycarotenoid dioxygenase (NCED), and carotenoid cleavage dioxygenases (CCD), have previously been cloned and characterized in sweetpotato (Kim et al., 2013; Kang et al., 2017; Kang et al., 2018; Suematsu et al., 2020). It is reported that silencing of CHYB, LCYB, or LCYE using RNAi increases carotenoid accumulation and resistance to abiotic stress such as heat, drought, and salt in transgenic sweetpotato plants and calli (Kim et al., 2012; Kim et al., 2013; Kim et al., 2014; Kang et al., 2017; Kang et al., 2018; Ke et al., 2019). Despite some genes and transcription factors related to carotenoid synthesis have been cloned in sweetpotato, the regulatory network controlling the variation in carotenoid content and composition is not well characterized, and rare regulators have been reported based on comparative transcriptome and co-expression network analysis.

In the present study, liquid chromatography tandem mass spectrometry (LC-MS/MS) was used to detect and quantify carotenoids in five sweetpotato cultivars with different flesh colors. At the same time, the differently expressed genes involved in carotenoid metabolism were identified by using transcriptional data in these five cultivars. In addition, co-expression gene modules were obtained, and key transcription

Abbreviations: GGPP, geranylgeranyl diphosphate; GGPS, pyrophosphatase synthase; IPP, isopentenyl diphosphate isomerase; DMAPP, dimethylallyl diphosphate; WGCNA, weight gene co-expression network analysis; RNA-Seq, transcriptome sequencing; LC-MS/MS, metabolomics; qRT-PCR, quantitative real-time PCR; DEGs, differentially expressed genes; DETFs, differentially expressed transcription factors; FDR, false discovery rate; PCC, pearson correlation coefficient; PSY, 15-cis-phytoene synthase; PDS, 15-cis-phytoene desaturase; Z-ISO, 15-cis- ζ -carotene isomerase; ZDS, ζ -carotene desaturase; CRTISO, polycopene isomerase; LCYB, lycopene β -cyclase; LCYE, lycopene ϵ -cyclase; DWARF27, β -carotene isomerase D27; CCD7, carotenoid cleavage dioxygenase 7; CHYB, β -carotene hydroxylase; CYP97B2, cytochrome P450 97B2-type epsilon-hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; NSY, neoxanthin synthase; NCED, 9-cis-epoxycarotenoid dioxygenase; ABA2, xanthoxin dehydrogenase; CYP707A, (+)-abscisic acid 8'-hydroxylase.

factors related to carotenoid metabolism were also screened through weighted gene co-expression network analysis (WGCNA). This work not only clarified the difference of carotenoid content and composition in tuberous roots of sweetpotato with different colors, but also provided important insights into the regulatory network of carotenoid metabolism in sweetpotato.

Materials and methods

Plant materials

Five sweetpotato cultivars with different flesh colors were used for metabolome and transcriptome analysis. The cultivars included, Shangshu 19 with white flesh (W), Okinawa-No.100 with light-yellow flesh (LY), Jieshu 95-16 with yellow flesh (Y), Guangshu 87 with orange flesh (O), and Pushu 32 with orange-red flesh (OR). All of the cultivars were planted using the standard production practices at the Baiyun experimental station (23°23'N, 113°26'E, 20 m above sea level) at Guangdong Academy of Agricultural Sciences, Guangzhou, China. Furthermore, samples were harvested in triplicates, 10 medium-sized tuberous roots from each cultivar were selected and divided into two parts. One part was used for color measurement, and the other part was cut into small pieces, mixed well, frozen in liquid nitrogen immediately and stored at -80°C for further metabolite extraction, transcriptome sequencing, and real-time PCR analysis, with three replicates for each sweetpotato cultivar.

Color measurement

The samples for color measurement were cut in half crosswise, then separately sliced into cross section with 1 cm thickness. The cross-sections of two pieces in the middle of each tuberous root were scanned using an A3 Unis Scanner (Uniscan® M1 Plus, UNIS, China), and the International Commission on Illumination CIE 1976 (L^* , a^* , b^*) color space values were measured using a software of Tomato Analyzer 3.0 (Rodríguez et al., 2010). Before measuring the color, the software was standardized with a color calibration card. The average of the three corresponding readings of each cultivar was considered and analyzed statistically as the final values. Results were described as L^* (brightness or lightness, positive towards white and negative towards black), a^* (red-green, positive towards red and negative towards green), b^* (yellow-blue, positive towards yellow and negative towards blue), H° (hue angle, calculated from the arctangent of b^*/a^* , 0° or 360° towards red, 90° towards yellow, 180° towards green, 270° towards blue), and C^* (chroma, calculated as $\sqrt{(a^{*2} + b^{*2})}$) (McGuire, 1992).

Carotenoid extraction

The carotenoid was extracted by the method described by Ma et al. (Ma et al., 2017) with slight modifications. The flesh samples of sweetpotato were freeze-dried in a freeze dryer (Xinzhi 16-0350, Ningbo, China) for 64 h with a minimum temperature of -50°C. Then lyophilized flesh samples were homogenized and powdered in a mill using a grinder with 30 Hz for 1 min (MM 400, Retsch, Germany). 50 mg freeze dried powder was extracted with a solvent containing n-hexane: acetone: ethanol (1:1:1, V/V/V), 0.01% BHT and internal standard were added. The extract was vortexed for 20 min at room temperature, centrifuged and the supernatant was collected. The residue was re-extracted repeating the above step. The extract was evaporated to dryness under nitrogen gas stream and re-suspended in mixed solution of methanol: methyltert-butyl ether (1:1, V/V). The solution was filtered through a 0.22 µm filter for further LC-MS analysis.

Identification and quantification of carotenoid

The sample extracts were measured using a UPLC-APCI-MS/MS system (UPLC, ExionLC™ AD, MS, Applied Biosystems 6500 Triple Quadrupole) with a YMC C30 column (3 µm, 100 mm × 2.0 mm). Mobile phases were MeCN-MeOH (3:1, V/V) containing 0.01% BHT and 0.1% formic acid (eluent A) and 100% MTBE containing 0.01% BHT (eluent B), and methanol was used as the probe wash. The flow rate was set to 0.8 mL/min, the column temperature was set at 28°C, and the injection volume was set as 2 µL. The gradient programs (eluent A: eluent B) were as follows: 100:0 V/V from 0 min, 30:70 V/V at 5.0 min, 5:95 V/V at 9.0 min, 100:0 V/V at 9.1 min, 100:0 V/V at 11.0 min (Inbaraj et al., 2008; Petry and Mercadante, 2017).

MS/MS detection was performed on AB 6500 Triple Quadrupole LC-MS/MS System, equipped with an atmospheric pressure chemical ionization (APCI) heated nebulizer interface. The APCI source operation parameters were as follows: ion source, APCI +; source temperature, 350°C; curtain gas (CUR) was set at 25.0 psi; Declustering potential (DP) and collision energy (CE) for individual MRM transition was done with further DP and CE optimization. A specific set of MRM transitions were monitored for each period according to the carotenoids eluted within this period.

The carotenoid analysis was performed using the Metware database (MWDB, Wuhan, China) constructed from the standards to qualitatively analyze the mass spectrometry data. We analyzed LC-MS/MS data from the carotenoids using Analyst 1.6.3 software (AB Sciex) with default parameters for automatic identification of changes and integrals in the MRM. Carotenoid retention times and ion pair information were used to modify chromatographic peaks for each carotenoid in

different samples (Figure S1). Standard curves of different carotenoids were drawn with the concentration ratio of external standard to internal standard as the abscissa and the peak area ratio of chromatographic peak as the ordinate (Table S1). Then the ratio of the integral peak area of each carotenoid detected in the samples to the peak area of the internal standard was substituted into the standard curves. Finally, the absolute content of carotenoids in the actual samples was calculated using the following formula: carotenoid content ($\mu\text{g/g}$) = $c \cdot V / 1000 / m$, where c is the concentration value obtained by substituting the integrated peak area ratio of the sample into the standard curve ($\mu\text{g/mL}$), V is the resuspension volume (μL), and m is the sample weight (g).

RNA-seq analysis and functional annotation

The total RNA was extracted from frozen samples of five sweetpotato genotypes (three replicates of each). RNA purity, concentration and integrity were analyzed on NanoPhotometer[®] spectrophotometer (IMPLEN, CA, USA), Qubit[®] RNA Assay Kit in Qubit[®] 2.0 Fluorometer (Life Technologies, CA, USA), and RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

One μg RNA of each sample was used for construction of sequencing libraries. Sequencing libraries were generated using NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] (NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. mRNA was purified from total RNA using poly-T oligo attached magnetic beads. The first and second strand cDNA was synthesized, double-strand cDNA was purified, the adapters were added. In order to select cDNA fragments of preferentially 250–300 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then sequencing libraries were sequenced on an Illumina NovaSeq 6000 system. The quality of sequence data was checked and adapters sequences and low-quality sequences were removed using Fastp with default parameters (Chen et al., 2018). Clean reads were mapped to the reference genome of sweetpotato (Yang et al., 2017) using HISAT2 (Kim et al., 2015). The gene alignment was calculated with FeatureCounts v1.6.2 (Liao et al., 2014), and the FPKM of each gene was calculated based on the gene length. Based on the original count data, differential gene expression between the samples was calculated using DESeq2 v1.22.1 (Varet et al., 2016). The genes with $|\log_2 \text{Fold Change}| \geq 1$ and False Discovery Rate (FDR) < 0.05 were called differentially expressed genes (DEGs). All transcriptome data have been uploaded to the China National Genomics Data Center (NGDC) database (BIG Sub - BioProject (cnb.ac.cn)), and the BioProject ID is PRJCA008295.

The gene enrichment analysis is performed based on the hypergeometric test. Gene function annotation was performed using four primary databases: Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) database, NCBI non-redundant protein sequences (NR) and Swiss-Prot.

Co-expression networks analysis

To identify the potential candidate genes involved in regulation of carotenoid accumulation, differentially expression genes (DEGs) and variability in carotenoid contents detected in the five cultivars were selected for performing integrative analysis. Weighted gene co-expression networks analysis (WGCNA) was performed in R package WGCNA (Langfelder and Horvath, 2008), based on 18,340 genes with normalized FPKM values and the carotenoid content in different samples. Cytoscape (Otasek et al., 2019) software was used to visualize co-expression network in each module.

Validation using quantitative real-time polymerase chain reaction

The total RNA was extracted from the flesh of five sweetpotato cultivars according to MiniBEST Plant RNA Extraction Kit (TaKaRa, Beijing, China). The first-strand cDNA was synthesized using TIANGEN FastKing gDNA Dispelling RT SuperMix (TIANGEN, Beijing, China). In this study, we randomly selected five genes from RNA-seq for qRT-PCR analysis, and used the *Ibactin* gene (AY905538) as a reference gene to correct gene expression. In addition, all primers used in this study were designed by Primer 3Plus (Untergasser et al., 2007), and listed in Supplement Table S2. Real-time quantitation PCR analysis was performed by Bio-RAD CFX96 system (Hercules, CA, United States) using a SsoFast[™] EvaGreen Supermix (Bio-Rad) at temperatures 95°C for 3 min and 40 cycles of 95°C for 15 s, 55°C for 10 s, and 72°C for 30 s. The melting curve, carrying out 61 cycles with 0.5°C increments from 65 to 95°C. The qRT-PCR data was generated in three replicates for each sample, and the quantitative data was analyzed by the $2^{-\Delta\Delta C_t}$ method.

Statistical analysis

All experiments were performed in three biological replicates in this study. The data of carotenoid contents were expressed as mean \pm standard deviations. Statistical significance and Pearson correlation coefficients were analyzed using SPSS v.26.0 (IBM, SPSS Inc., Chicago, IL, USA). Significant differences among five cultivars were calculated using one-way ANOVA

followed by Duncan's multiple way test at $P < 0.05$ level. The heatmaps of gene expression level and Venn diagram were drawn by TBtools (Chen et al., 2020).

Results

Color assessment

Five sweetpotato cultivars with different flesh colors: white (W), light-yellow (LY), yellow (Y), orange (O), and orange-red (OR) were used in this study (Figure 1A). To understand the color difference between these five cultivars, the color parameters (L^* , a^* , b^* , H° , and C^*) of fleshs were measured and illustrated in Figure 1B. The parameter values showed significant variation among five sweetpotato cultivars. However, the parameters a^* , b^* , and C^* exhibited a similar tendency across the cultivars, contrary to the trend of parameter L^* . Interestingly, we observed that L^* value decreased, and a^* , b^* and C^* values increased with the change of flesh color gradient from white to orange-red. The parameter H° of OR cultivar was the lowest, tending to orange-red, while LY cultivar was the

largest, tending to yellow. These results were consistent with visual perception. Flesh color in sweetpotato is one of the most important factors which influences the flavor and aroma and therefore attracts consumer attention in the market. Currently, orange-red sweetpotatoes are more popular in the market, which may be largely related to their rich color.

Composition and content of carotenoids

We estimated the content and composition of carotenoid in five sweetpotatoes of different flesh colors by UPLC-MS/MS (Table 1). A total of 14 carotenoids were identified in the non-saponified extracts from five sweetpotato cultivars, of which 11 were identified as free carotenoids or intermediates involved in the carotenoid biosynthetic pathway (including 3 carotenes and 8 xanthophylls) and 3 were identified as carotenoid monoesters. Especially, a total of 9 carotenes namely β -carotene, violaxanthin palmitate, antheraxanthin, zeaxanthin, violaxanthin, neoxanthin, lutein, β -cryptoxanthin, apocarotenal were detected in all five cultivars. Zeaxanthin palmitate was detected in LY, Y, O and OR cultivars, violaxanthin myristate was

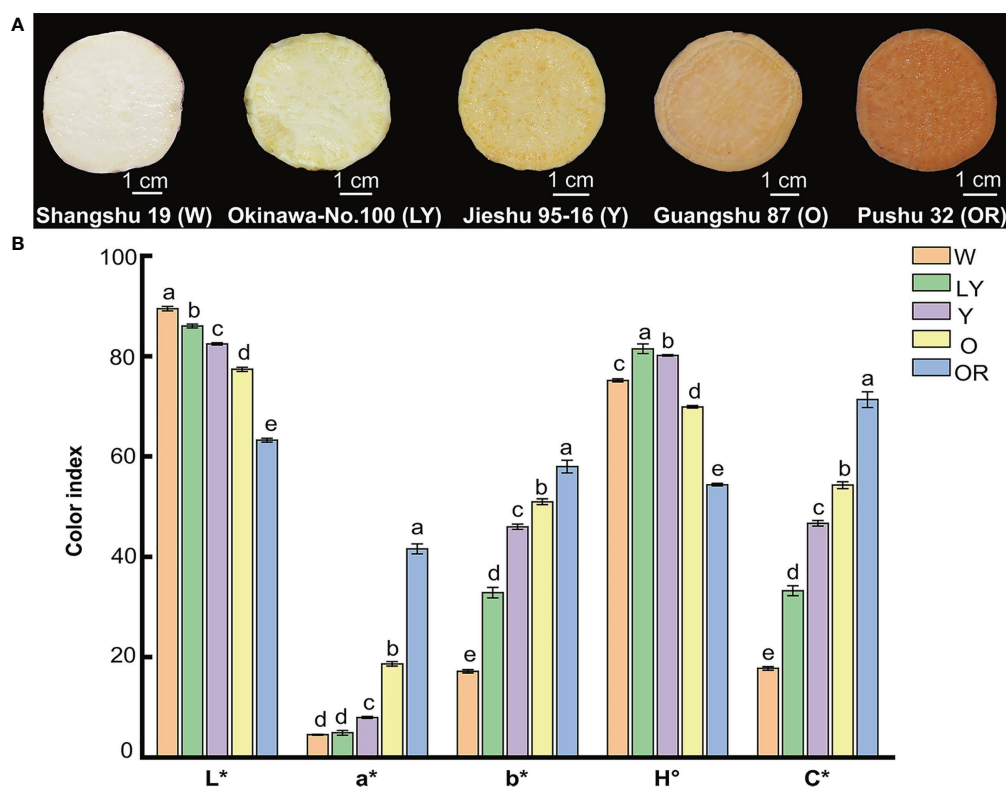


FIGURE 1
Phenotypic characteristics and color parameters of five sweetpotato fleshs. (A) Flesh color of five cultivars. Shangshu 19, Okinawa-No.100, Jieshu 95-16, Guangshu 87, Pushu 32 cultivars were renamed as W, LY, Y, O, OR, respectively. (B) The color indexes (L^* , a^* , b^* , H° , C^*) of five sweetpotato fleshs. Lowercase letters a, b, c, d and e represent significance at $p \leq 0.05$. Error bars represent mean \pm SD (n = 3).

TABLE 1 Content of carotenoids in five sweetpotato cultivars.

Class	Carotenoid ($\mu\text{g/g DW}$)	W	LY	Y	O	OR	P value	CV (%)
Carotenes	γ -Carotene	–	–	–	$0.28 \pm 0.02\text{b}$	$2.90 \pm 0.53\text{a}$	< 0.001	187.51
	β -Carotene	$0.35 \pm 0.02\text{c}$	$0.67 \pm 0.11\text{c}$	$3.12 \pm 0.30\text{c}$	$85.07 \pm 5.40\text{b}$	$385.33 \pm 29.14\text{a}$	< 0.001	162.71
	(E/Z)-Phytoene	–	–	–	$2.50 \pm 0.20\text{b}$	$10.03 \pm 0.81\text{a}$	< 0.001	160.98
Carotenoid esters	Violaxanthin myristate	–	–	$0.09 \pm 0.01\text{b}$	$0.22 \pm 0.04\text{a}$	$0.19 \pm 0.01\text{a}$	< 0.001	100.00
	Violaxanthin palmitate	$0.54 \pm 0.07\text{c}$	$0.72 \pm 0.05\text{c}$	$0.51 \pm 0.04\text{c}$	$1.53 \pm 0.12\text{b}$	$2.69 \pm 0.42\text{a}$	< 0.001	73.33
	Zeaxanthin palmitate	–	$0.22 \pm 0.02\text{c}$	$1.04 \pm 0.10\text{b}$	$4.84 \pm 0.36\text{a}$	$0.93 \pm 0.12\text{b}$	< 0.001	129.79
Xanthophylls	Antheraxanthin	$0.20 \pm 0.03\text{d}$	$0.97 \pm 0.05\text{b}$	$0.77 \pm 0.07\text{c}$	$1.24 \pm 0.05\text{a}$	$1.10 \pm 0.14\text{b}$	< 0.001	44.71
	Zeaxanthin	$0.26 \pm 0.01\text{d}$	$5.12 \pm 0.53\text{c}$	$9.29 \pm 1.84\text{b}$	$8.84 \pm 0.62\text{b}$	$11.63 \pm 1.07\text{a}$	< 0.001	59.89
	Violaxanthin	$0.51 \pm 0.04\text{c}$	$0.46 \pm 0.04\text{c}$	$0.24 \pm 0.02\text{c}$	$2.65 \pm 0.20\text{b}$	$3.37 \pm 0.82\text{a}$	< 0.001	95.83
	Neoxanthin	$0.26 \pm 0.03\text{b}$	$0.23 \pm 0.03\text{b}$	$0.09 \pm 0.01\text{c}$	$0.62 \pm 0.10\text{a}$	$0.72 \pm 0.09\text{a}$	< 0.001	66.67
	Lutein	$3.34 \pm 0.05\text{b}$	$1.19 \pm 0.04\text{c}$	$1.07 \pm 0.10\text{c}$	$1.43 \pm 0.07\text{c}$	$9.28 \pm 1.08\text{a}$	< 0.001	100.00
	β -Cryptoxanthin	$0.67 \pm 0.06\text{c}$	$2.10 \pm 0.17\text{c}$	$11.23 \pm 1.04\text{c}$	$35.90 \pm 1.71\text{b}$	$139.00 \pm 13.75\text{a}$	< 0.001	143.59
	Apocarotenal	$0.01 \pm 0.00\text{c}$	$0.01 \pm 0.00\text{c}$	$0.01 \pm 0.01\text{c}$	$0.05 \pm 0.01\text{b}$	$0.28 \pm 0.05\text{a}$	< 0.001	157.14
	Echinenone	–	–	–	$0.02 \pm 0.00\text{b}$	$0.16 \pm 0.02\text{a}$	< 0.001	150.00

For each carotenoid, different lowercase letters a, b, c, and d in a row indicated significances among means ($P < 0.05$); SD means standard deviation; CV means coefficient of variance.

detected in Y, O and OR cultivars, whereas, γ -carotene, (E/Z)-phytoene and echinenone were detected only in O and OR cultivars.

The content of 14 carotenoids showed significant variation among the five sweetpotato cultivars ($P < 0.001$). OR cultivar showed highest content of carotenoids among all cultivars except violaxanthin myristate, zeaxanthin palmitate, and antheraxanthin. With the change of color gradient from white to orange-red, the content of β -carotene and β -cryptoxanthin gradually increased. Among the 14 carotenoids, β -carotene was found to be dominant in OR and O cultivars, $385.33 \mu\text{g/g}$ in OR cultivar and $85.07 \mu\text{g/g}$ in O cultivar. β -cryptoxanthin and zeaxanthin were enriched in Y and LY cultivars, respectively, while the highest level of carotenoid in W cultivar was lutein ($3.34 \mu\text{g/g}$). These results indicated that abundant yellow xanthophylls (β -cryptoxanthin, zeaxanthin and lutein) and orange carotenoids (especially β -carotene) may give the yellow and orange colors in sweetpotato flesh.

Transcriptomic analysis and function annotation

After filtering the adapter sequences and low quality reads, the average clean reads of 46.26 million, 50.32 million, 46.71 million, 43.39 million and 43.93 million were obtained from W, LY, Y, O, and OR colored flesh, respectively. The average GC contents were $44.09 \sim 46.21\%$, and the average Q30 values were $93.58 \sim 94.31\%$ among five cultivars (Table S3). All clean reads were mapped to the sweetpotato reference genome, and the mapping rate was ranged from 74.54% to 83.54%. Pearson correlation coefficients were calculated between samples based on the number of fragments per kilobase of exon per million fragments mapped (FPKM) values. The correlation coefficients

between the biological replicates of the same cultivar were higher than 0.8 (Figure 2A), indicating good reproducibility of the biological repeats. Overall, these results indicated that the quantity and quality of the sequencing data were suitable for the downstream analysis.

A total of 18,340 differentially expressed genes were identified by pairwise comparison between five cultivars based on $|\log_2 \text{Fold Change}| \geq 1$ and $\text{FDR} < 0.05$ (Table S4). The clustering heatmap of all DEGs was analyzed (Figure 2B). A total of 5,142 in LY cultivar, 4,399 in Y cultivar, 5,618 in O cultivar and 6,276 in OR cultivar when compared with W cultivar, and 1,025 DEGs were identified as the common across these combinations (Figure 2C). Among these, 2,440, 1,740, 2,252 and 2,744 DEGs were up-regulated in LY vs. W, Y vs. W, O vs. W and OR vs. W combinations, respectively (Tables S5–S8; Figure 2D).

KEGG pathway analysis showed that a total of 1,637, 1,534, 1,788 and 2,030 DEGs from the combination of LY vs. W, Y vs. W, O vs. W, OR vs. W participated in a total of 130, 131, 131 and 127 KEGG pathways, respectively (Table S9 and Figure S2). Among these pathways, terpenoid backbone biosynthesis and photosynthesis - antenna proteins were significantly upregulated in LY vs. W, Y vs. W, O vs. W combinations. The pathways associated with circadian rhythm-plant and isoflavonoid biosynthesis were highly upregulated in O vs. W, OR vs. W combinations, while the carotenoid biosynthesis pathway highly enriched in O vs. W combination. Based on the GO database, the DEGs were classified into 55, 51, 55, and 57 subcategories from LY vs. W, Y vs. W, O vs. W, OR vs. W, respectively (Table S10 and Figure S3). In GO enrichment analysis we observed that secondary metabolite biosynthetic pathways were active in LY/Y/OR vs. W combinations. The above results suggested that the circadian rhythm-plant pathway, light system and secondary

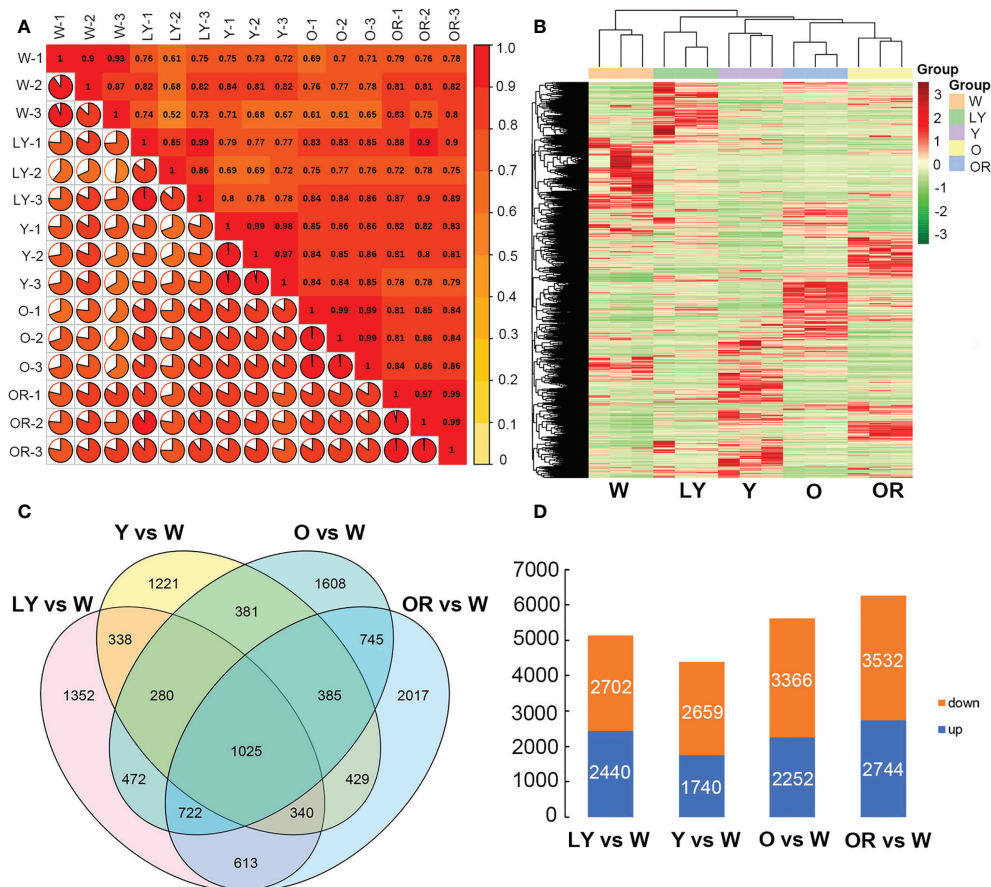


FIGURE 2

Gene expression profiling of five samples. (A) Correlation coefficients between gene expression data sets from three biological duplicates. (B) Heatmap illustrates expression pattern of differentially expressed genes in five sweetpotato fleshes. (C) Venn diagram showing differentially expressed genes between each pairwise comparison. (D) Bar graphs showing the upregulated and downregulated genes in four pairwise combinations.

metabolite biosynthesis might be the key factors responsible for accumulation of pigments in sweetpotato flesh.

DEGs involved in carotenoid metabolic pathway

The flesh color in sweetpotatoes not solely depends on carotenoid content, it is also regulated by the expression level of the intermediate key genes involved in carotenoid biosynthesis and degradation. Further to reveal the mechanism of carotenoid variation in sweetpotato flesh, the DEGs involved in carotenoid metabolic pathway were analyzed. KEGG enrichment analysis and NR annotation showed that, there are 27 structural genes participated in carotenoid metabolic pathway among the four combinations (LY vs. W, Y vs. W, O vs. W, OR vs. W), and the sweetpotato carotenoid metabolic pathway were mapped (Figure 3). The pathway was constructed based on KEGG and references (Suematsu et al., 2020; Xu et al., 2021; Yuan et al., 2021; KEGG,

2022). Of which 10 DEGs namely *Or* (1), *CRTISO* (1), *CYP97B2* (1), *CHYB* (1), *ZEP* (2), *NSY* (1) and *DWARF27* (3) were playing major role in carotenoid biosynthesis pathway and 17 DEGs namely *CCD1* (8), *CCD4* (1), *CCD7* (1), *NCED* (4) and *CYP707A* (3) were related to carotenoid degradation pathway. The expression of 27 genes was shown in Tables S11, S12. Except for *CYP707A*-2 gene, the expression levels of these genes varied significantly among five cultivars. Heatmap (Figure 3) illustrates the expression patterns of 27 DEGs. For qRT-PCR analysis, 5 key DEGs were randomly selected to validate their expression level (Table S13).

Pearson correlation analysis between the gene expression and the carotenoid contents was shown in Table S14. The expression levels of genes involved in biosynthesis and degradation of various carotenoids showed significant correlation with carotenoid content. For instance, *DWARF27-1* gene expression was significantly positively correlated with all individual carotenoids. The expression of *CCD1*-2 gene was significantly negatively correlated with all individual carotenoids except for lutein and zeaxanthin palmitate. The expression of one *ZEP* gene, g1103 showed significantly negative

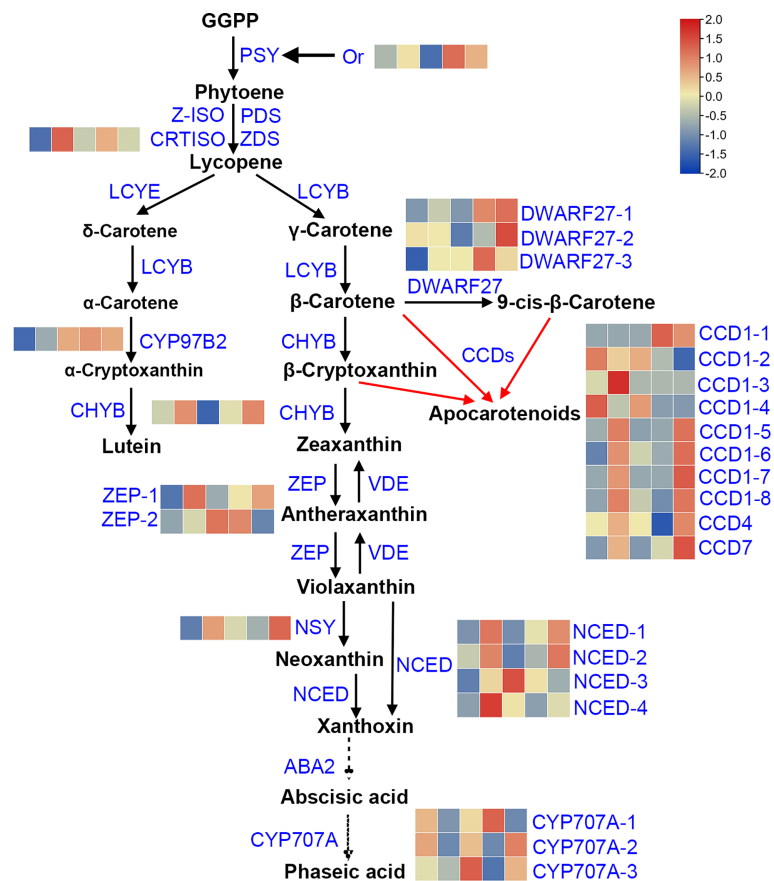


FIGURE 3

Pathway analysis of carotenoid metabolic in sweetpotato flesh. This pathway was constructed based on the KEGG pathway and references. GGPP, geranylgeranyl diphosphate; PSY, 15-cis-phytoene synthase; PDS, 15-cis-phytoene desaturase; Z-ISO, 15-cis-ζ-carotene isomerase; ZDS, ζ-carotene desaturase; CRTISO, prolycopene isomerase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; DWARF27, β-carotene isomerase D27; CCD1, carotenoid cleavage dioxygenase 1; CCD4, carotenoid cleavage dioxygenase 4; CCD7, carotenoid cleavage dioxygenase 7; CHYB, β-carotene hydroxylase; CYP97B2, cytochrome P450 97B2-type epsilon-hydroxylase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; NSY, neoxanthin synthase; NCED, 9-cis-epoxycarotenoid dioxygenase; ABA2, xanthoxin dehydrogenase; CYP707A, (+)-absciscic acid 8'-hydroxylase. Heatmap of the expression of DEGs involved in carotenoid synthesis pathway, rows and columns represent gene names and samples (from left to right is W, LY, Y, O and OR), respectively.

correlation with lutein content, and the expression of the other *ZEP* gene g1106 showed significantly positive correlation with antheraxanthin content. The expression of *CCD1-6/7/8*, *CCD7*, *CHYB*, *DWARF27-2*, *NCED-2* and *NSY* genes was significantly positively correlated with the content of three carotenes. *Or* gene expression was significantly positively correlated with the content of three carotenoid esters, antheraxanthin, violaxanthin and neoxanthin. However, the expression of *NCED-1*, *NCED-4*, *CYP707A-2*, *CYP707A-3* and *CRTISO* was not significantly correlated with the content of any carotenoid compositions.

WGCNA module analysis

WGCNA is a system biology approach to describe association patterns of gene expression between different samples and can be

used for finding the highly correlated genes related to phenotypes (Langfelder and Horvath, 2008). In this study, after removing 90% of the DEGs with FPKM values < 1, a total of 9,151 DEGs were retained and divided into 12 modules (Figure 4A and Table S15). The clustering heatmap of DEGs in gene modules was drawn to show the correlations between gene modules (Figure S4). The analysis of module-carotenoid relationships revealed that the blue and brown co-expressed modules were significantly associated with carotenoid metabolites (Figure 4B). The blue module containing 1,077 DEGs exhibited a highly negative correlation with carotenoid contents except for zeaxanthin with correlation coefficient ($r \leq -0.5$ and $P < 0.05$), and the brown module containing 1,056 DEGs had a highly positive correlation with the content of antheraxanthin, violaxanthin myristate, zeaxanthin, zeaxanthin palmitate, violaxanthin, neoxanthin, and violaxanthin palmitate.

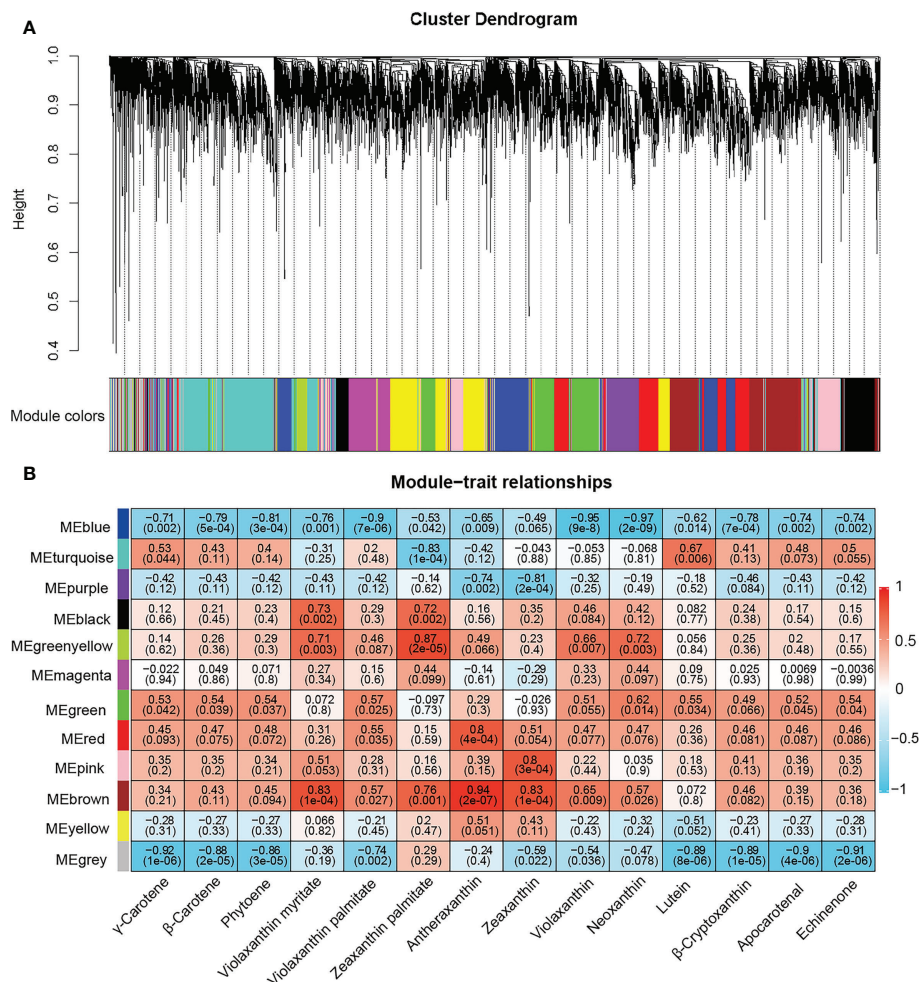


FIGURE 4

Weighted gene co-expression network analysis (WGCNA) of differentially expressed genes. (A) Hierarchical clustering tree showing 12 gene modules of co-expressed genes, each module labeled with different color. Each leaf in branch of the tree represented one gene. (B) Relationship between gene module and carotenoids contents in sweetpotato cultivars. Each row corresponds to a module, labeled with different colors. Each column represents a carotenoid component. The values in each cell represent the correlation coefficient and *P* value between the module and carotenoids and displayed according to the color scale on the right.

Heatmaps for DEGs in blue and brown modules showed two distinct clusters in each module (Figure S5). In blue module, the genes of cluster I showed higher expression levels in OR and O cultivars than in W, LY and Y cultivars. However, the genes of cluster II showed lower expression levels in OR and O cultivars than in W, LY and Y cultivars. Furthermore, the genes of cluster I in the brown module were highly expressed in W cultivar and the genes of cluster II showed higher expression levels in O cultivar. Therefore, we concluded that the genes in blue and brown modules must be playing different roles in the carotenoid metabolic pathways.

Co-expression network analysis identified carotenoid related TFs

To further identify transcription factors related to carotenoid content and composition and to determine the relationship between gene modules, a correlation network was constructed with edge weight ≥ 0.1 in blue and brown modules. When gene significance (GS) and eigengene connectivity (kME) was set as ≥ 0.9 or ≤ -0.9 , 34 and 51 genes were used to construct a co-expression network, 5 and 10 transcription factors were identified relating to carotenoid accumulation in blue and brown modules, respectively

(Figures 5A, B and Table S16). In the blue module, NAC transcription factor family *NAC72* and *NAC86* interact with *ERF* family member *ERFABR1* and *GATA* family member *ATHB-13*. The transcription factors *NAC72*, *ERFABR1* and *PRR73* were highly expressed in Y, and lowly expressed in OR and O cultivars, while the transcription factors *NAC86* and *ATHB-13* showed higher expression abundance in O and OR cultivars (Figure S6). In the brown network, transcription factor *SR45a* showed interaction with *CCCH20*, *PIF1*, *ATHB-15*, *HSF30*, *NF-YA-10*, *TCP8* and *ARR11*. Furthermore, the expression trend of *SR45a* was completely consistent with the expression pattern of *ARR11*, *TCP8*, *PIF1*, and *ATHB-15* among the five sweetpotato cultivars. Interestingly, *MYB3* interacts with *bZIP11*, but the expression pattern of the two genes was opposite. For instance, *MYB3* was highly expressed in W, while the expression of *bZIP11* was lower in W cultivar (Figure S6).

Additionally, we constructed the networks between differentially expressed transcription factors (DETFs) and DEGs related to carotenoid metabolic pathway in the blue and brown modules and carotenoids content based on GS to further

investigate the role of hub genes in regulation of carotenoid accumulation (Figures 5C, D and Table S17). In the blue module, *NAC72* and *ERFABR1* were significantly negatively correlated with the content of neoxanthin and violaxanthin, while *NAC86* was positively correlated with these two carotenoids. In case of brown module, except for *MYB3*, the expression of DETFs and DEGs were significantly positively correlated with zeaxanthin or antheraxanthin content. These results indicated that not only the genes from blue and brown modules are correlated with the content of carotenoid, but also there are some important transcription factors that regulate the biosynthesis and accumulation of carotenoid in sweetpotato flesh.

Validation of the expression of key DEGs by qRT-PCR

To validate the expression level of key genes related to carotenoid accumulation, we randomly selected 5 DEGs

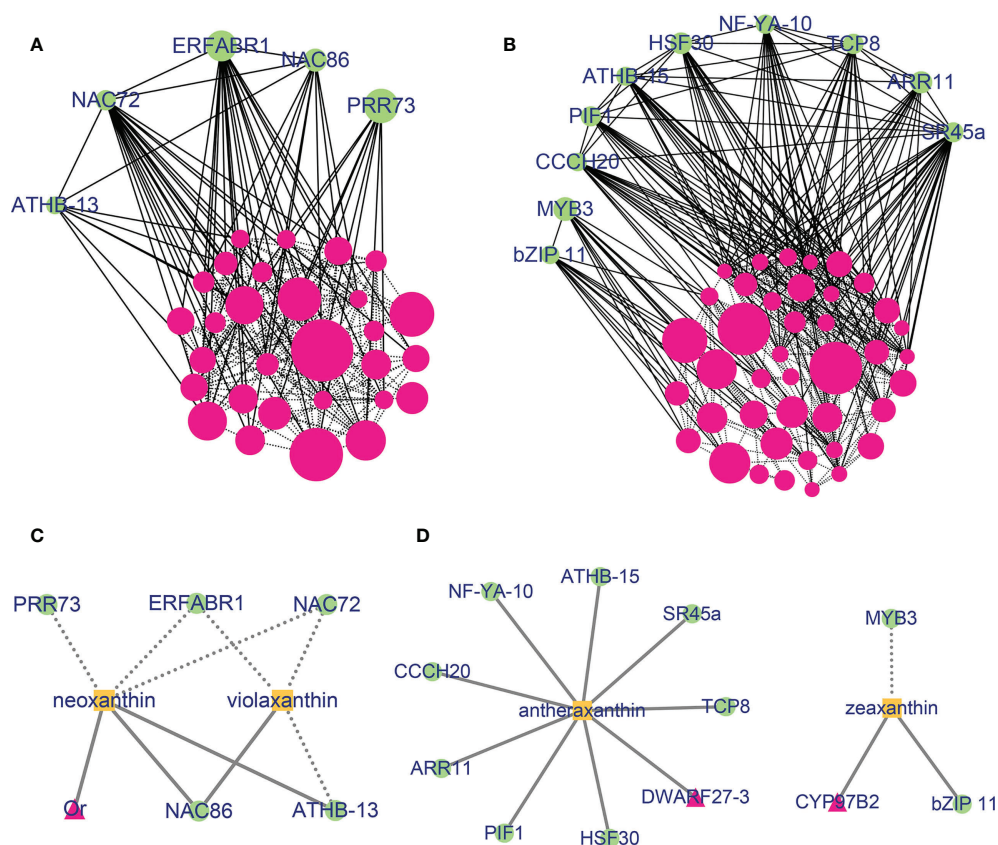


FIGURE 5

Co-expression network of genes in the blue and brown modules. (A) Gene co-expression network of the blue module. (B) Gene co-expression network of the brown module. The solid lines mean the interaction with transcription factors and dotted lines represent interaction between two genes. Dot sizes represent the weight between genes. Co-expression network of carotenoid metabolism genes and transcription factor in the blue (C) and brown (D) modules with carotenoids. The solid lines present that the genes were positively with carotenoids (GS > 0) and dotted lines represent that the genes were negatively with carotenoids (GS < 0). Rose red, proteins; Green, transcription factors; Orange, carotenoids.

namely *CHYB*, *ZEP*, *NSY*, *NCED* and *CYP707A* for qRT-PCR. The expression of *ZEP-1*, *NCED-3* genes was higher in LY, OR, O and Y than in W cultivars, which was consistent with the result of RNA-seq. A correlation analysis exhibited a significant correlation between qRT-PCR and RNA-seq with correlation coefficient > 0.8 (Table S13 and Figure S7). These results indicated that our transcriptome data is reliable.

Discussion

Carotenoids in different flesh-colored sweetpotatoes

Many flowers, fruits, seeds and roots exhibit various colors such as white, yellow, orange to red, which are mainly determined by the different proportions of carotenoid content (HOWITT and POGSON, 2006; Zhou et al., 2020; Xu et al., 2021). The correlation analysis between color index and carotenoid contents showed that L^* and H^o values were significantly negatively correlation with carotenoids contents, while a^* , b^* , and C^* values were opposite (Table S18). The results showed that the color in tuberous root of sweetpotato, ranging from white, light-yellow, yellow, orange to orange-red, is also attributed to the different level and ratio of carotenoids. In this study, 14 carotenoids were identified in five sweetpotato cultivars with different flesh colors by LC-MS/MS (Table 1). The individual carotenoids contents were significantly different among five sweetpotato cultivars. The content of β -carotene and β -cryptoxanthin in W cultivar was lower, and gradually increased with the deepening of flesh color from white to orange-red, which was consistent with the findings of Alam et al. (Alam et al., 2016) and Islam et al. (Islam et al., 2016). However, β -cryptoxanthin and zeaxanthin were main carotenoids in yellow and light-yellow cultivars, β -carotene present in high proportion in orange and orange-red cultivars, which was consistent with Ishiguro's report (Ishiguro et al., 2010). Our results indicated that the yellow and orange flesh of sweetpotatoes may be caused by the yellow xanthophylls and orange β -carotene carotenoids.

Carotenoids encompass carotenes and xanthophylls (also called carotene derivatives that contain one or more oxygen atoms) (Armstrong, 1997). What's more, xanthophylls possessing hydroxyl groups are present in the form of free or esterified xanthophylls with different fatty acids (Mariutti and Mercadante, 2018). In present study, zeaxanthin palmitate, violaxanthin palmitate and violaxanthin myristate were identified in small amounts for the first time in sweetpotato flesh (Table 1). Similar to the finding of Breithaupt and Bamedi (Breithaupt and Bamedi, 2002), violaxanthin and zeaxanthin were the main esterified carotenoids in potato flesh, mainly esterified to the fatty acid of palmitic acid and myristic acid. Fernandez-Orozco et al. (Fernandez-Orozco et al., 2013) found that the esterified carotenoid

fraction was positively correlated with the total carotenoid content in sixty potato cultivars. However, that was not consistent with the results of this study, which may be due to the low lipid content and carotenoid esters content in sweetpotatoes. These results suggest that the carotenoid esterification may be species-specific and further experiments are needed to verify. Carotenoids are present in the form of esterification in plants, which influences the interaction between carotenoids and other molecules and affects carotenoid sequestration and accumulation (Watkins and Pogson, 2020). In this study, the content of esterified carotenoid is very low, so it is not necessary to be considered in the study of the mechanism of carotenoid accumulation in sweetpotatoes.

The key genes involved in carotenoid metabolic pathway

The mechanism of carotenoid accumulation is mainly regulated by biosynthesis, degradation, and sequestration processes (Watkins and Pogson, 2020). Currently, researchers studying carotenoid accumulation are mainly concerned with transcriptional regulation, especially those genes involved in carotenoid metabolic pathway. Studies have shown that *CRTISO* (polycopene isomerase) is one of the enzymes that catalyze the synthesis of lycopene from phytoene, and has a critical role in carotenoid biosynthesis pathway in the dark and in non-photosynthetic tissues (Isaacson et al., 2002). Suematsu et al. (Suematsu et al., 2020) identified *ZEP* paralog (g1103.t1) as a key gene involves in carotenoid accumulation by regulating epoxidation of β -carotene and β -cryptoxanthin in yellow-fleshed sweetpotato. Cytochrome P450 (*CYP97*) family is the key enzyme that catalyzes carotene to produce xanthophylls (Li and Wei 2020). In Arabidopsis, it was reported that *CYP97C1* hydroxylated both β - and ϵ -rings of α -carotene, while *CYP97A3* was most active on β -ring of α -carotene. Subsequent studies revealed that *CYP97B* family members were β -ring specific, similar to *CYP97A* (Niu et al., 2020). In this study, *CYP97B2* gene (g33351) encodes carotenoid epsilon-hydroxylase was identified in O vs. W and OR vs. W combinations. The function annotation of *CYP97B2* gene in sweetpotato was not consistent with the research results of Niu et al. (Niu et al., 2020), this may be due to the different carotenoid accumulation mechanism of sweetpotato and Arabidopsis. *Or* gene encodes an orange protein involved in accumulation of carotenoid in plants. Overexpression of *Or* gene has been shown to increase carotenoid content by post-transcriptionally regulating PSY (Zhou et al., 2015). In the present study, we found that the expression of *Or* gene (g20467) was significantly different between five sweetpotato cultivars, but the PSY gene showed no significant difference. This may be attributed to that PSY gene expression is not only affected by orange protein but also regulated by the other transcription factors, such as phytochrome-interaction factors (PIFs) (Toledo-Ortiz et al., 2010).

Interestingly, 27 DEGs were identified involved in the carotenoid metabolic pathways. Among these, except for *CRTISO*, *NCED-1/4* and *CYP707A-2/3*, the gene expression levels were significantly correlated with carotenoid content (Table S14). However, the variation of carotenoid accumulation among five sweetpotato cultivars was not completely consistent with its upstream carotenogenic genes expression trend. The *CRTISO* gene (g59407) showed different expression patterns in five sweetpotato cultivars, and that expression level was upregulated in LY and O cultivars (Figure 3), lycopene was not detected in any of the sweetpotato cultivars. The content of neoxanthin was higher in OR and O than in W, LY and Y cultivars, while the expression of *NSY* gene was significantly higher in OR and LY than in Y, O and W cultivars. Although violaxanthin and antheraxanthin were highly accumulated in OR and O cultivars, the expression level of *ZEP* gene was higher in LY and Y cultivars. We concluded that the accumulation of carotenoid was not directly correlated with the expression of upstream carotenogenic genes. These results are consistent with the findings of Watkins et al. (Watkins and Pogson, 2020). Therefore, the study on the mechanism of carotenoid coloration should focus on the regulation of transcription factors as well as gene transcription in carotenoid metabolic pathway.

Transcription factor related to carotenoid metabolic pathway in sweetpotato

In recent years, several transcription factors that regulate and control the carotenoid metabolic pathway have been identified in several crop species (Ampomah-Dwamena et al., 2019; Yuan et al., 2021). In present study, we performed WGCNA to identify transcription factors that coordinately regulate the biosynthesis and accumulation of carotenoids. We found 15 differentially expressed transcription factors that were related to carotenoid content in sweetpotato fleshes. As an underground part of the plant, the tuberous root of sweetpotato is susceptible to abiotic stresses, such as water, temperature, salt and other environmental cues (Stanley and Yuan, 2019). Previous studies showed that some transcription factors were positively or negatively correlated with carotenogenic genes, and regulated the expression of abiotic stress-responsive genes in crops. The CCH-type zinc finger protein *IbC3H18* is a key regulator of abiotic stress response in sweetpotato, such as ABA signaling, reactive oxygen species scavenging (Zhang et al., 2019). In *Arabidopsis*, the serine/arginine-rich splicing factor, *SR45a*, was found to participate in the regulation of salinity tolerance (Li et al., 2021). Huang et al. reported that the heat-stress factor *HSFA6b* participated in the positive regulation of ABA mediated salt and drought resistance (Huang et al., 2016). We found that *CCCH20*, *HSF30* and *SR45a* genes showed different expression in five sweetpotato fleshes, and were significantly positively correlated with antheraxanthin content. The

downstream products of antheraxanthin, violaxanthin and neoxanthin, can be degraded by *NCED* to produce abscisic acid (ABA). We speculated that abiotic stress response of sweetpotato regulated the biosynthesis of carotenoid and these transcription factors might regulate the biosynthesis of carotenoid downstream products by participating through ABA signal transduction.

Phytochrome-interacting factors 1 (*PIF1*) specifically inhibited *PSY* gene by interacting with other transcription factors to regulate carotenoid synthesis (Toledo-Ortiz et al., 2010). In kiwifruit, a *R2R3-MYB* transcription factor, *MYB7* was found to regulate carotenoid accumulation by activating the promoter of the lycopene beta-cyclase (*LCYB*) gene (Ampomah-Dwamena et al., 2019). In this study, the expression of *PIF1* gene was positively correlated with the content of antheraxanthin, and the *MYB3* was negatively related to zeaxanthin content. The results indicated that *PIF1* and *MYB3* transcription factors may be activate or inhibit the expression of carotenogenic genes to regulate the biosynthesis of antheraxanthin and zeaxanthin in sweetpotatoes. Although the key genes involved in the carotenoid biosynthesis pathway and the transcription factors related to carotenoid metabolism have been identified, the regulatory mechanisms and gene functions of carotenoid coloration still need to be further investigated and verified.

The carotenoid metabolites in five sweetpotato cultivars with different flesh colors were successfully determined by LC-MS/MS. A total of 14 carotene metabolites were detected in sweetpotato fleshes, and β -carotene was obviously dominant in orange and orange-red colored fleshes, β -cryptoxanthin and zeaxanthin were enriched in yellow and light-yellow colored fleshes, while lutein was abundant in white colored flesh. Moreover, in the present study, zeaxanthin palmitate, violaxanthin palmitate and violaxanthin myristate were identified in small amounts for the first time in sweetpotato fleshes. Combined with different metabolites and transcriptome analysis, 27 enzymatic genes in carotenoid biosynthesis and degradation were identified in four sweetpotato combinations. Through WGCNA, 15 differentially expressed transcription factors were found to be significantly correlated with carotenoids content, which may be involved in carotenoid accumulation and affect the formation of sweetpotato flesh color.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

Conceptualization and methodology: RJ and ZW; material provider: BJ; formal analysis: RJ and RZ; data curation: RJ; writing original draft preparation: RJ; writing review and editing: ZW, GL, RZ,

SG, CT, and BJ; visualization: RJ; funding acquisition: ZW. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by National Key R&D Program of China (nos. 2019YFD1000700 and 2019YFD1000701), the earmarked fund for CARS-10-Sweetpotato and the Guangdong Modern Agro-industry Technology Research System (2022KJ111).

Conflict of interest

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.

References

- Alam, M. K., Rana, Z. H., and Islam, S. N. (2016). Comparison of the proximate composition, total carotenoids and total polyphenol content of nine orange-fleshed sweet potato varieties grown in bangladesh. *Foods* 5 (3), 64. doi: 10.3390/foods5030064
- Ampomah-Dwamena, C., Thrimawithana, A. H., Dejnopratt, S., Lewis, D., Espley, R. V., and Allan, A. C. (2019). A kiwifruit (*Actinidia deliciosa*) R2R3-MYB transcription factor modulates chlorophyll and carotenoid accumulation. *New Phytol.* 221 (1), 309–325. doi: 10.1111/nph.15362
- Armstrong, G. A. (1997). GENETICS of EUBACTERIAL CAROTENOID BIOSYNTHESIS: A colorful tale. *Annu. Rev. Microbiol.* 51 (1), 629–659. doi: 10.1146/annurev.micro.51.1.629
- Beck, G., Coman, D., Herren, E., Ruiz-Sola, M. Á., Rodríguez-Concepción, M., Grussem, W., et al. (2013). Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.* 82 (4–5), 393–416. doi: 10.1007/s11103-013-0070-z
- Brethaupt, D. E., and Bamedi, A. (2002). Carotenoids and carotenoid esters in potatoes (*Solanum tuberosum* L.): new insights into an ancient vegetable. *J. Agr. Food Chem.* 50 (24), 7175–7181. doi: 10.1021/jf0257953
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34 (17), i884–i890. doi: 10.1093/bioinformatics/bty560
- CIP (2020). CIP Annual Report 2019. *Discovery to Impact: Science-based solutions for global challenges* (pyrmont, Australia: CIP), 9. doi: 10.4160/02566311/2019
- Eggersdorfer, M., and Wyss, A. (2018). Carotenoids in human nutrition and health. *Arch. Biochem. Biophys.* 652, 18–26. doi: 10.1016/j.abb.2018.06.001
- Fernandez-Orozco, R., Gallardo-Guerrero, L., and Hornero-Méndez, D. (2013). Carotenoid profiling in tubers of different potato (*Solanum* sp) cultivars: Accumulation of carotenoids mediated by xanthophyll esterification. *Food Chem.* 141 (3), 2864–2872. doi: 10.1016/j.foodchem.2013.05.016
- Fraser, P. D., and Bramley, P. M. (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* 43 (3), 228–265. doi: 10.1016/j.plipres.2003.10.002
- HOWITT, C. A., and POGSON, B. J. (2006). Carotenoid accumulation and function in seeds and non-green tissues. *Plant Cell environment.* 29 (3), 435–445. doi: 10.1111/j.1365-3040.2005.01492.x
- Huang, Y., Niu, C., Yang, C., and Jinn, T. (2016). The heat-stress factor HSF6b connects ABA signaling and ABA-mediated heat responses. *Plant Physiol.* 172 (2), 860–2016. doi: 10.1104/pp.16.00860
- Inbaraj, B. S., Lu, H., Hung, C. F., Wu, W. B., and Lin, C. L. (2008). Determination of carotenoids and their esters in fruits of *Lycium barbarum* Linnaeus by HPLC–DAD–APCI–MS. *J. Pharm. Biomed. Anal.* 47, 812–818. doi: 10.1016/j.jpba.2008.04.001
- Isaacson, T., Ronen, G., Zamir, D., and Hirschberg, J. (2002). Cloning of tangerine from tomato reveals a carotenoid isomerase essential for the production of β -carotene and xanthophylls in plants. *Plant Cell* 14 (2), 333–342. doi: 10.1105/tpc.010303
- Ishiguro, K., Yoshinaga, M., Kai, Y., Maoka, T., and Yoshimoto, M. (2010). Composition, content and antioxidative activity of the carotenoids in yellow-fleshed sweetpotato (*Ipomoea batatas* L.). *Breed. Sci.* 60 (4), 324–329. doi: 10.1270/jsbbs.60.324
- Islam, S. N., Nusrat, T., Begum, P., and Ahsan, M. (2016). Carotenoids and β -carotene in orange fleshed sweet potato: A possible solution to vitamin A deficiency. *Food Chem.* 199, 628–631. doi: 10.1016/j.foodchem.2015.12.057
- Kang, L., Park, S., Ji, C. Y., Kim, H. S., Lee, H., and Kwak, S. (2017). Metabolic engineering of carotenoids in transgenic sweetpotato. *Breed. Sci.* 67 (1), 27–34. doi: 10.1270/jsbbs.16118
- Kang, C., Zhai, H., Xue, L., Zhao, N., He, S., and Liu, Q. (2018). A lycopene β -cyclase gene, *lbcYB2*, enhances carotenoid contents and abiotic stress tolerance in transgenic sweetpotato. *Plant Sci.* 272, 243–254. doi: 10.1016/j.plantsci.2018.05.005
- KEGG (2022). Available at: <https://www.kegg.jp/pathway/map00906>.
- Ke, Q., Kang, L., Kim, H. S., Xie, T., Liu, C., Ji, C. Y., et al. (2019). Down-regulation of lycopene ϵ -cyclase expression in transgenic sweetpotato plants increases the carotenoid content and tolerance to abiotic stress. *Plant Sci.* 281, 52–60. doi: 10.1016/j.plantsci.2019.01.002
- Kim, S. H., Ahn, Y. O., Ahn, M., Lee, H., and Kwak, S. (2012). Down-regulation of β -carotene hydroxylase increases β -carotene and total carotenoids enhancing salt stress tolerance in transgenic cultured cells of sweetpotato. *Phytochem. (Oxford)* 74, 69–78. doi: 10.1016/j.phytochem.2011.11.003
- Kim, S. H., Kim, Y., Ahn, Y. O., Ahn, M., Jeong, J. C., Lee, H., et al. (2013). Downregulation of the lycopene ϵ -cyclase gene increases carotenoid synthesis via the β -branch-specific pathway and enhances salt-stress tolerance in sweetpotato transgenic calli. *Physiol. Plantarum.* 147 (4), 432–442. doi: 10.1111/j.1399-3054.2012.01688.x
- Kim, H. W., Kim, J. B., Poovan, S., Chung, M. N., Cho, S. M., Lee, Y. M., et al. (2014). Effect of processing conditions on the content of *cis/trans* carotene isomers as provitamin A carotenoids in Korean sweet potato varieties. *Int. J. Food Sci. Nutr.* 65 (7), 821–826. doi: 10.3109/09637486.2013.854742
- Kim, D., Langmead, B., and Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* 12 (4), 357–360. doi: 10.1038/nmeth.3317

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.993682/full#supplementary-material>

- Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinf.* 9 (1), 559. doi: 10.1186/1471-2105-9-559
- Liao, Y., Smyth, G. K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30 (7), 923–930. doi: 10.1093/bioinformatics/btt656
- Li, Y., Guo, Q., Liu, P., Huang, J., Zhang, S., Yang, G., et al. (2021). Dual roles of the serine/arginine-rich splicing factor SR45a in promoting and interacting with nuclear cap-binding complex to modulate the salt-stress response in arabidopsis. *New Phytol.* 230 (2), 641–655. doi: 10.1111/nph.17175
- Liu, Y., Lv, J., Liu, Z., Wang, J., Yang, B., Chen, W., et al. (2020). Integrative analysis of metabolome and transcriptome reveals the mechanism of color formation in pepper fruit (*Capsicum annuum* L.). *Food Chem.* 306, 125629. doi: 10.1016/j.foodchem.2019.125629
- Li, Y., and Wei, K. (2020). Comparative functional genomics analysis of cytochrome P450 gene superfamily in wheat and maize. *BMC Plant Biol.* 20 (1), 93. doi: 10.1186/s12870-020-2288-7
- Mariutti, L. R. B., and Mercadante, A. Z. (2018). Carotenoid esters analysis and occurrence: What do we know so far? *Arch. Biochem. Biophys.* 648, 36–43. doi: 10.1016/j.abb.2018.04.005
- Ma, G., Zhang, L., Iida, K., Madono, Y., Yungyuen, W., Yahata, M., et al. (2017). Identification and quantitative analysis of beta-cryptoxanthin and beta-citraurin esters in Satsuma mandarin fruit during the ripening process. *Food Chem.* 234, 356–364. doi: 10.1016/j.foodchem.2017.05.015
- Mcguire, R. G. (1992). Reporting of objective color measurements. *Hortscience* 27 (12), 1254–1255. doi: 10.21273/HORTSCI.27.12.1254
- Niu, G., Guo, Q., Wang, J., Zhao, S., He, Y., and Liu, L. (2020). Structural basis for plant lutein biosynthesis from α -carotene. *Proc. Natl. Acad. Sci.* 117 (25), 14150–14157. doi: 10.1073/pnas.2001806117
- Otasek, D., Morris, J. H., Bouças, J., Pico, A. R., and Demchak, B. (2019). Cytoscape automation: Empowering workflow-based network analysis. *Genome Biol.* 20 (1), 185. doi: 10.1186/s13059-019-1758-4
- Petry, F., and Mercadante, A. (2017) New method for carotenoid extraction and analysis by HPLC-DAD-MS/MS in freeze-dried citrus and mango pulps. *J. Braz. Chem. Society* 29, 205–215. doi: 10.21577/0103-5053.20170127
- Rodríguez, G. R., Moysenko, J. B., Robbins, M. D., Huarachi Morejón, N., Francis, D. M., and van der Knaap, E. (2010). Tomato analyzer: A useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *J. Visualized Experiments.* 37, 1856. doi: 10.3791/1856
- Sajilata, M. G., Singhal, R. S., and Kamat, M. Y. (2008). The carotenoid pigment zeaxanthin—a review. *Compr. Rev. Food Sci. F.* 7 (1), 29–49. doi: 10.1111/j.1541-4337.2007.00028.x
- Stanley, L., and Yuan, Y. W. (2019). Transcriptional regulation of carotenoid biosynthesis in plants: So many regulators, so little consensus. *Front. Plant Sci.* 101017. doi: 10.3389/fpls.2019.01017
- Suematsu, K., Tanaka, M., Kurata, R., and Kai, Y. (2020). Comparative transcriptome analysis implied a ZEP paralog was a key gene involved in carotenoid accumulation in yellow-fleshed sweetpotato. *Sci. Rep.-UK.* 10 (1), 20607. doi: 10.1038/s41598-020-77293-7
- Toledo-Ortiz, G., Huq, E., and Rodríguez-Concepción, M. (2010). Direct regulation of phytoene synthase gene expression and carotenoid biosynthesis by phytochrome-interacting factors. *Proc. Natl. Acad. Sci.* 107 (25), 11626–11631. doi: 10.1073/pnas.0914428107
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., and Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* 35, W71–W74. doi: 10.1093/nar/gkm306
- Varet, H., Brillet-Gueguen, L., Coppee, J. Y., and Dillies, M. A. (2016). SARTools: A DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-seq data. *PLoS One* 11 (6), e157022. doi: 10.1371/journal.pone.0157022
- Watkins, J. L., and Pogson, B. J. (2020). Prospects for carotenoid biofortification targeting retention and catabolism. *Trends Plant Sci.* 25 (5), 501–512. doi: 10.1016/j.tplants.2019.12.021
- World Food Prize Foundation (2016). Available at: https://www.worldfoodprize.org/en/laureates/20102019_laureates/2016_andrade_bouis_low_and_mwanga/.
- Xu, X., Lu, X., Tang, Z., Zhang, X., Lei, F., Hou, L., et al. (2021). Combined analysis of carotenoid metabolites and the transcriptome to reveal the molecular mechanism underlying fruit colouration in zucchini (*Cucurbita pepo* L.). *Food Chemistry: Mol. Sci.* 2, 100021. doi: 10.1016/j.fochms.2021.100021
- Yang, J., Moeinzadeh, M., Kuhl, H., Helmuth, J., Xiao, P., Haas, S., et al. (2017). Haplotype-resolved sweet potato genome traces back its hexaploidization history. *Nat. Plants* 3, 696–703. doi: 10.1038/s41477-017-0002-z
- Yuan, P., Umer, M. J., He, N., Zhao, S., Lu, X., Zhu, H., et al. (2021). Transcriptome regulation of carotenoids in five flesh-colored watermelons (*Citrullus lanatus*). *BMC Plant Biol.* 21 (1), 203. doi: 10.1186/s12870-021-02965-z
- Zhang, H., Gao, X., Zhi, Y., Li, X., Zhang, Q., Niu, J., et al. (2019). A non-tandem CCCH-type zinc-finger protein, IbC3H18, functions as a nuclear transcriptional activator and enhances abiotic stress tolerance in sweet potato. *New Phytol.* 223 (4), 1918–1936. doi: 10.1111/nph.15925
- Zhou, W., Niu, Y., Ding, X., Zhao, S., Li, Y., Fan, G., et al. (2020). Analysis of carotenoid content and diversity in apricots (*Prunus armeniaca* L.) grown in China. *Food Chem.* 330, 127223. doi: 10.1016/j.foodchem.2020.127223
- Zhou, X., Welsch, R., Yang, Y., Álvarez, D., Riediger, M., Yuan, H., et al. (2015). Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proc. Natl. Acad. Sci.* 112 (11), 3558–3563. doi: 10.1073/pnas.1420831112



OPEN ACCESS

EDITED BY

Xia Xin,
Chinese Academy of Agricultural
Sciences, China

REVIEWED BY

Lixia Wang,
Institute of Crop Sciences, Chinese
Academy of Agricultural Sciences,
China
Bikash Adhikari,
Mississippi State University,
United States

*CORRESPONDENCE

Harsh Nayyar
harshnayyar@hotmail.com
Uday Chand Jha
u9811981@gmail.com
Kadambot H. M. Siddique
kadambot.siddique@uwa.edu.au

SPECIALTY SECTION

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

RECEIVED 13 September 2022

ACCEPTED 07 November 2022

PUBLISHED 21 November 2022

CITATION

Chaudhary S, Jha UC, Paul PJ,
Prasad PVV, Sharma KD, Kumar S,
Gupta DS, Sharma P, Singh S,
Siddique KHM and Nayyar H (2022)
Assessing the heat sensitivity of
Urdbean (*Vigna mungo* L. Hepper)
genotypes involving physiological,
reproductive and yield traits under
field and controlled environment.
Front. Plant Sci. 13:1042999.
doi: 10.3389/fpls.2022.1042999

COPYRIGHT

© 2022 Chaudhary, Jha, Paul, Prasad,
Sharma, Kumar, Gupta, Sharma, Singh,
Siddique and Nayyar. 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(s) 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.

Assessing the heat sensitivity of Urdbean (*Vigna mungo* L. Hepper) genotypes involving physiological, reproductive and yield traits under field and controlled environment

Shikha Chaudhary¹, Uday Chand Jha^{2*}, Pronob J. Paul³,
P. V. Vara Prasad⁴, Kamal Dev Sharma⁵, Sanjeev Kumar⁶,
Debjyoti Sen Gupta², Parul Sharma⁷, Sarvjeet Singh⁷,
Kadambot H. M. Siddique^{8*} and Harsh Nayyar^{1*}

¹Department of Botany, Panjab University, Chandigarh, India, ²Crop Improvement Division, Indian Council of Agricultural Research (ICAR)-Indian Institute of Pulses Research, Kanpur, India,

³International Rice Research Institute, South-Asia Hub, Hyderabad, India, ⁴Sustainable Intensification Innovation Lab, Kansas State University, Manhattan, KS, United States, ⁵Department of Agricultural Biotechnology, Choudhary Sarwan Kumar (CSK) Himachal Pradesh Agricultural University, Palampur, India, ⁶Department of Plant Sciences, Central University of Punjab, Bhatinda, India, ⁷Department of Plant Breeding and Genetics, Punjab Agricultural University, Ludhiana, India, ⁸The University of Western Australia (UWA) Institute of Agriculture, The University of Western Australia, Perth, WA, Australia

The rising temperatures are seriously impacting the food crops, including urdbean; hence efforts are needed to identify the sources of heat tolerance in such crops to ensure global food security. In the present study, urdbean genotypes were evaluated for heat tolerance under natural outdoor for two consecutive years (2018, 2019) and subsequently in the controlled environment of the growth chamber to identify high temperature tolerant lines. The genotypes were assessed involving few physiological traits (membrane damage, chlorophyll, photosynthetic efficiency, stomatal conductance, lipid peroxidation), reproductive traits (pollen germination % and pollen viability %) and yield related traits (total number of pods plant⁻¹, total seeds plant⁻¹, single seed weight and seed yield plant⁻¹). Based upon these tested traits, PantU31, Mash114, UTTARA and IPU18-04 genotypes were identified as promising genotypes for both years under heat stress condition. Further confirming heat tolerance, all these four tolerant and four sensitive genotypes were tested under controlled environment under growth chamber condition. All these four genotypes PantU31, Mash114, UTTARA and IPU18-04 showed high chlorophyll content, photosynthetic efficiency, stomatal conductance, leaf area, pods plant⁻¹, total seeds plant⁻¹ and low reduction in pollen germination % and pollen viability under stress heat stress condition. Moreover, yield and yield related traits viz., pods plant⁻¹, seeds plant⁻¹, single seed weight and seed yield plant⁻¹ showed very strong positive correlation with

pollen germination and pollen viability except electrolyte leakage and malondialdehyde content. Thus, these genotypes could be potentially used as donors for transferring heat tolerance trait to the elite yet heat-sensitive urdbean cultivars.

KEYWORDS

urdbean, heat stress, genetic variability, physiological traits, biochemical traits

1 Introduction

Due to global climate change projections, heat waves are predicted to expand in many regions of the world imposing a huge threat to the agricultural security (Riaz et al., 2021). Variability around the optimum temperature is going to surge that will affect the complete life cycle or phenology of the plant (Chaturvedi et al., 2021). Continuously rising temperature (heat stress) has wide range of impacts on the overall morphology, anatomy and physiology of the plants (Chen et al., 2014). At sub-cellular levels, these impacts can be assessed using various biochemical and molecular approaches. Although heat stress has the potential to affect all the stages of plant but some stages are more vulnerable to heat stress; reproductive or seed filling stages are highly affected due to heat stress (Allakhverdiev et al., 2008; Hedhly, 2011). Moreover, the effects of high temperature are plant species- and stage- specific, the severity further depends on the duration and intensity of stress (Li et al., 2018). Various reports have suggested that heat stress disturbs the morphology of the plant by reducing its plant height, leaf area and root architecture (Chaudhary et al., 2020). At the cellular level, heat stress leads to protein denaturation, enzyme inactivation, membrane damage, exaggerate ROS generation, loss in water status and cellular viability. In leaves, photosynthetic machinery is reported as most sensitive to high temperature (Wahid et al., 2007; Bitá and Gerats, 2013). Loss of chlorophyll content, denaturation of D1 protein of photosystem II and reduced carbon metabolism are mainly responsible for reducing photosynthetic rate (Allakhverdiev et al., 2008). Of the reproductive organ development stages; male gametophyte development and seed filling processes are reported to be extremely sensitive to even a few degree rise in temperature that can result in substantial yield loss (Hedhly, 2011). Impaired transport of sucrose to the developing reproductive organs under heat stress may restrict the flower development that brings out more flower abortion, pod abortion and shrivelled seeds (Bhandari et al., 2016). Plant responds to such damages by reprogramming and activating various mechanisms related to production of antioxidants, phytohormones, osmolytes, primary and secondary metabolites to ensure their survival (Chebrolov

et al., 2016; Sharma et al., 2016; Jha et al., 2022). Therefore, correlation of crop phenology with temperature fluctuations is crucial for the better understanding of the impacts and defence strategies employed by plants for its adaptation.

Urdbean (*Vigna mungo* L. Hepper) is an important summer season food legume, cultivated mostly in many tropical and sub-tropical countries of Asia, Africa, America, and Australia (Joshi and Rao, 2017). Optimum temperature for its proper growth and development is 25–35°C and being a temperature sensitive crop, its yield is drastically reduced under high temperature exceeding 35°C (Anitha et al., 2016; Sen Gupta et al., 2021). Very little reports are available about the heat stress impacts as well as defence responses, especially at reproductive stage of this crop (Sen Gupta et al., 2021). It is vital to identify and characterise heat tolerant urdbean genotypes as well as to find out some leaf and pollen-based traits and mechanisms underlying heat tolerance. Heat tolerant urdbean genotypes can increase the cultivation of this food legume in summer season as well at warmer locations to extend its cultivation status. Thus, the aim of the present study was to screen selective number of genotypes of urdbean to heat stress in 2 successive years under outdoor environment to identify heat tolerant genotypes, followed by their validation and characterisation under controlled environment of the growth chamber involving some leaf and pollen-based traits.

2 Methodology

2.1 Field and growth chamber experiments

Urdbean genotypes (26) were procured from Indian Institute of Pulse Research, Kanpur, India and Punjab Agricultural University, Ludhiana, India (Supplementary Table S1). These genotypes were assessed for their heat tolerance under outdoor environment and controlled conditions of the growth chamber at the Department of Botany, Panjab University, Chandigarh, India. Urdbean seeds were raised in pots (8L capacity) containing a mixture of soil, sand, farmyard manure

[2:1:1 (v/v)] and Tri-calcium phosphate fertilizer 10 mg kg⁻¹. Seeds were soaked in distilled water overnight (12 h) and subsequently inoculated with suitable strain of *Rhizobia* before sowing. There were 5 pots genotype⁻¹ and each pot had 5 seeds that were thinned to 3 plants pot⁻¹ after emergence for their proper growth. Plants were fully irrigated daily (twice; morning and evening) to avoid any water paucity. Plants were arranged in a randomized complete block design. Meteorological data (daily temperature and relative humidity) from date of sowing to harvesting was recorded throughout the entire cropping season (Supplementary Figure S1). To evaluate the effects of heat stress against the control temperature, crop was sown twice during cropping season and for two subsequent years (Summers of 2018, 2019) (i) during the normal conditions (control), in the last week of March 2018, when the day/night temperatures (<35/25°C) were optimum for the plant's growth and ii) in the last week of April, to expose the plants to heat stress (>40/30°C). (Supplementary Figure S1, Supplementary Table S2). The plants after harvest were recorded for number of pods, seeds, seed yield plant⁻¹ and single seed weight.

For validation of the results, a subsequent study was conducted in the growth chamber under the controlled conditions on some selected contrasting genotypes (4 heat-tolerant and 4 heat-sensitive genotypes, 5 pots genotype⁻¹ having 2 plants pot⁻¹). These plants were initially raised in the outdoor natural environment to achieve full vegetative growth (Average temperature <35/25°C; average RH 61/41%; Max/min) and were subsequently transferred to growth chamber at the onset of bud stage for further analysis. To avoid any kind of heat shock situation, temperature was gradually raised (2°C per day) up to 42/32°C. The plants were maintained at this temperature up to maturity. Simultaneously, the control plants were maintained at 35/25°C.

After 10d of heat stress, fully expanded leaves at 2nd and 3rd positions from the topmost youngest leaf were from control and heat-stressed environments were evaluated for various physiological traits viz. SPAD chlorophyll content, chlorophyll fluorescence (Fv/Fm), electrolyte leakage (to assess membrane damage), stomatal conductance, leaf area, relative leaf water content, and malondialdehyde (MDA). The reproductive traits (pollen viability and pollen germination) were tested from flowers after 5d exposure to heat stress. All these traits were further correlated with yield traits like total number of pods plant⁻¹, total number of seeds plant⁻¹, total seed yield plant⁻¹, single seed weight.

2.2 Physiological, reproductive and yield traits

To assess the effects of heat stress on the plant growth and yield, various traits were studied; data were taken from three plants in triplicates genotype⁻¹, pooled and averaged. Mean values of replicates are presented through tables and figures.

2.3 Physiological traits

2.3.1 Chlorophyll content

Chlorophyll content (SPAD value) was measured using Apogee-SPAD meter and its readings were taken between 10.00 and 11.00 h of a fully expanded tagged leaf on alternative days at full vegetative and reproductive stage from 30 DAS (days after sowing) (Devi et al., 2022).

2.3.2 Chlorophyll fluorescence

PS II activity/stability or photosynthetic efficiency was measured as chlorophyll fluorescence. Readings were taken between 10.00-11.00 h of a fully expanded leaf using the dark adapted test of modulated chlorophyll fluorometer OS1-FL (Opti-Sciences, Tyngsboro, MA, United States) (Sharma et al., 2016).

2.3.3 Electrolyte leakage

Stress injury to leaves was measured as electrolyte leakage. Fresh leaves (1.0 g) were collected and washed three times with deionised water to remove surface adhering electrolytes. Plant tissue was placed in closed vials containing 10 ml of deionised water and incubated it for 25°C on a rotary shaker for 24 h; the electrical conductivity of the solution (L₁) was checked using a conductivity meter (ELICO CM 180, Hyderabad, India). Then the final conductivity (L₂) was measured after heating samples in a water bath at 120°C for 20 min (Lutts et al., 1996). Electrolyte leakage was calculated as (L₁/L₂) × 100. The electrolyte leakage was expressed as electrical conductivity in μmhos g⁻¹ DW.

2.3.4 Stomatal conductance

Stomatal conductance was measured from a fully expanded leaf using a portable leaf porometer (model SC1 Decagon Devices, Pullman, WA, United States) at 11.00 h and was expressed as m molm⁻²s⁻¹ (Awasthi et al., 2014).

2.3.5 Leaf area

Area of fully expanded tagged leaves was determined using a measurement scale and multiplied with a 'leaf factor' (method derived from urdbean from the ratio of actual and measured leaf area of many types of leaves from top to bottom of a plant) (Sharma et al., 2016).

2.3.6 Relative leaf water content

RLWC was measured by the method of Barrs and Weatherley (1962). Fresh leaves were collected and were washed three times to remove any kind of debris. After drying with blotters, they were weighed (fresh weight, FW) and then floated in the distilled water in a petri dish. After 2 h, leaves were taken out of petri dish, reweighed and surface dried with blotters. Leaves were then oven-dried at 110°C for 24h and again weighed for dry weight (DW). Final values for relative leaf water content was calculated as (FW-DW)/(TW-DW) × 100.

2.3.7 Malondialdehyde content

Lipid peroxidation of the cell membrane was measured as malondialdehyde (MDA) content (Heath and Packer, 1968). One hundred mg fresh leaf tissue was extracted in 10 mL of 0.1% trichloroacetic acid (TCA). The homogenate was centrifuged at 15,000 rpm for 5 min. Supernatant was used as extract. Afterward, 4 mL of 0.5% thiobarbituric acid (in 20% trichloroacetic acid) was added to a 1-ml of the supernatant. This mixture was heated at 95°C for 30 minutes followed by immediate cooling in ice bath. Re-centrifugation of this mixture was performed again at 10,000 rpm for 10 min and the absorbance of the supernatant was taken at 532 nm. Values were expressed as nmol g⁻¹ DW.

2.4 Reproductive traits

For evaluating reproductive function, flowers were collected 5 days after exposure to heat stress and assessed for following traits.

2.4.1 Pollen germination

For testing pollen germination, pollen grain samples were taken in three replicates and each replicate consisted of five flowers genotype⁻¹ (Brewbaker and Kwack, 1963). Pollen grains were collected and immersed in few drops of pollen germinating medium (10% sucrose, 990 mM potassium nitrate (pH 6.5), 1.64 mM boric acid, 812 mM magnesium sulphate and 1.3 mM calcium nitrate) (Kaushal et al., 2013).

2.4.2 Pollen viability

Around 100 pollen grains were tested for the pollen viability with 0.5% acetocarmine/Alexander stain per genotype in three replicates. Selection of viable pollen grains was made on the basis of size (fully expanded), shape (triangular or spherical) and concentration of stain taken by them. Pollen grains were collected from freshly opened flowers and were pooled and checked for their viability (Kaushal et al., 2013).

2.5 Yield traits

For obtaining yield data, three plants genotype⁻¹ in three replications (9 plants genotype⁻¹) were harvested at maturity, wrapped in paper bags and dried in an oven at 65°C for at least three days. After drying, the total number of pods and seeds, total seed weight and single seed weight plant⁻¹ were calculated (Sharma et al., 2016).

2.6 Statistical analyses

Urdbean plants were grown in outdoor environment for 2 consecutive years as well as under controlled environment of the

growth chamber using RCBD. The analysis of data for computing standard errors and least significant differences ($P < 0.05$) was performed using 2-factorial (temperature × genotypes) design using OPSTAT statistical software (CCS, HAU, Hisar, India). Genotypic correlation, heritability were analysed by using GenStat 15 software. The Euclidean dissimilarity matrix was constructed involving all the genotypes and traits, and were clustered using Ward's method (Patterson and Thompson, 1971). The principal component analysis was done using the R package factoextra, and heat map analysis was performed according to Babicki et al. (2016).

3 Results

3.1 Physiological traits

3.1.1 Electrolyte leakage

Electrolyte leakage (EL%) is one of the important physiological traits measuring membrane damage used for screening heat stress tolerant genotypes in plants. Heat stress significantly ($P < 0.01$) damaged the membranes (Supplementary Figure S2, Supplementary Table S3). EL increased by 49 and 51% in heat-stressed plants, compared to controls, in the first and second years, respectively. Based on this trait, Mash 114 (18.5%, 17.73%), PantU31(21.77%, 20.8%), UTTARA (21.43%, 20.73%), IPU18-04 (18.17%, 19.73%) genotypes revealed low value for EL % under heat stress environment for both years. However, the genotypes IPU 18-6 (25.13%, 26.9%), Mash 218 (26.47%, 26%), SuG1153 (26.23%, 26.9%) exhibited high value for EL% under heat stress environment for both years suggesting their heat stress sensitivity. The high heritability values (82.6% and 86.85, for first and second years, respectively) for this trait were noted under heat stress see Table 1.

3.1.2 Stomatal conductance

Stomatal conductance (gS) varied significantly ($P < 0.01$) across the genotypes in plants exposed to high temperature (Supplementary Figure S3, Supplementary Table S3). As a result of high temperature, gS decreased by 12 and 15% over control in 1st and 2nd year, respectively. Under heat stress environment, Mash 114 (45.6, 40.47) and Pant G 31(43.9, 46.53) genotypes showed high value for stomatal conductance in both years. Regarding heritability for gS, 96.5% (during the first year) and 95.2% (during second year) heritability values were noted (Table 1).

3.1.3 Chlorophyll content

A significant genetic variation ($P < 0.01$) was noticed in chlorophyll content among the genotypes under heat stress (Supplementary Figure S4, Supplementary Table S3). The range of leaf chlorophyll content was noted to be 11.4-19.8 mg g⁻¹ FW during the first year and 12.3- 21.3 mg g⁻¹ FW during

TABLE 1 General statistics of various traits in urdbean genotypes under heat stress environment.

	Heritability	CV%	Mean	Range
Heat stress 1 st year				
Chlorophyll content	85.5	8.1	15.1	11.4-19.8
chlorophyll fluorescence	90.1	5.3	0.53	0.41-0.61
Electrolyte leakage%	82.6	7.3	24.2	18.2-28
Leaf area	83.7	6.6	18.4	13.9-21.4
Stomatal conductance	96.5	8.1	28.9	22.3-45.6
Pods plant ⁻¹	98.5	13.8	5.72	2.4-15
Seeds plant ⁻¹	99.4	11.2	21.03	6.2-62
Seed yield plant ⁻¹	99.2	19	0.64	0.11-2.7
Single seed weight	92.2	12.1	0.03	0.02-0.04
Heat stress 2 nd year				
Chlorophyll content	92.7	7.6	15.7	12.3-21.3
Chlorophyll fluorescence	97.3	2.9	0.54	0.44-0.66
Electrolyte leakage%	86.8	6.7	24.4	17.3-27.7
Leaf area	89.8	6.1	17.6	14.7-20.6
Pods plant ⁻¹	99	11.3	6.16	2.8-15.9
Stomatal conductance	95.2	8.6	30	25.5-46.5
Seeds plant ⁻¹	99.6	9.4	22	5.3-71
Seed yield plant ⁻¹	99.6	15	0.73	0.16-2.03
Single seed weight	94.5	10.1	0.03	0.02-0.04
Growth chamber heat stress				
Chlorophyll content	94.5	7.5	16.35	13.27-19.67
chlorophyll fluorescence	98.8	3.7	0.54	0.42-0.66
Electrolyte leakage%	96.2	4.5	24.5	20.27-28.03
Leaf area	94.5	7.5	16.35	13.27- 19.67
Malondialdehyde	98.8	5.3	27.75	19.3-33.97
Pollen germination %	99.4	6.8	36.7	15.3-54.17
Pollen viability%	98.8	6.7	43.24	24.5-61.20
Relative water content	99.3	2.9	63.66	49.57- 79.1
Stomatal conductance	99.1	5.5	25.03	14.37-33.7
Seeds plant ⁻¹	99.4	9.1	29	4-52
Seed yield plant ⁻¹	99.5	10.6	1.26	0.15- 2.38
Single seed weight	98.7	7.4	0.03	0.02-0.04
Pods plant ⁻¹	99.4	9.1	8.79	2.7-14.73

second year under heat stress environment. An average reduction of 22 and 30% was observed due to heat stress, relative to controls, in 1st and 2nd year, respectively. High value for chlorophyll content was observed in Mash 114 (19.67, 18.97), PantG31 (19.77, 20.47), UTTARA (17, 20.4) genotypes under heat stress environment for both the years. This trait also showed high heritability (85.5% and 92.7% for 1st and 2nd years, respectively) and could be vital for selecting heat tolerant urdbean lines (Table 1).

3.1.4 Chlorophyll fluorescence

Significant genetic variability for chlorophyll fluorescence (ChlF) (Fv/Fm) was noted under heat stress environment ($P < 0.01$) (Supplementary Figure S5, Supplementary Table S3).

Mean value for Fv/Fm was noted to be 0.53 during the first year and 0.54 during second year under heat stress environment. Heat stress caused about 28% reduction over control in both the years. The genotypes Mash 114 (0.61, 0.66), PantU31 (0.61, 0.65), UTTARA (0.6, 0.65) showed high value for ChlF under heat stress for both years. Heritability for this trait was noted to be 90.1% and 97.3% during the first year and second year, respectively (Table 1).

3.1.5 Leaf area

Significant genetic variation ($P < 0.01$) was noted in leaf area (LA) among the tested genotypes under hot environment for both years (Supplementary Figure S6, Supplementary Table S3). LA decreased by 23 and 28% in heat-stressed plants, over

controls, in the 1st and 2nd years, respectively. Substantial genetic variability for this trait was noted under heat stress environment ranging from 13.9–21.4 cm² (during the first year) and 14.7–20.6 cm² (during second year). High heritability (83.7%) recorded during the first year and 89.8% during second year, suggested that this trait could be used for screening heat tolerance in urdbean (Table 1).

3.2 Yield and yield-related traits

Significant ($P < 0.01$) genetic variability for pods plant⁻¹ (Figure 1) seeds plant⁻¹ (Figure 2), seed yield plant⁻¹ (Figure 3) and single seed weight (Figure 4) were recorded under heat stress environment for both years (Supplementary Table S3).

Under high temperature, maximum pods plant⁻¹ decreased by 82.4 and 83.7%, maximum seeds plant⁻¹ by 94.7 and 94.3%, maximum seed yield plant⁻¹ by 91.5 and 95.2% and single seed weight by 26 and 32% over the respective controls in 1st and 2nd year, respectively. The UTTARA genotype retained highest pod number plant⁻¹ (15, 16) followed by PantU31(15,15), IPU18-04 (14,16) under heat stress environment for both years. For seeds

plant⁻¹ trait, Mash114 (63, 66), UTTARA (63, 71) and IPU18-04 (62, 63) showed promising results under heat stress environment for both years. Likewise, Mash114 (47.7%, 41.6%), IPU18-04 (43.34%, 48.9%), UTTARA (57.39%, 38%), and PantU31 (55.45%, 28.7%) showed lower reduction percentage for seed yield plant⁻¹ for both years, and thus could be highly heat tolerant genotypes. Under heat stress environment, high heritability with 98.5%, 99.4% and 99.2% was noted for pods plant⁻¹, seeds plant⁻¹ and seed yield plant⁻¹, respectively during the first year. Similarly, these traits showed high heritability under hot environment during second year also (Table 1).

3.3 Validation of selected heat tolerant and heat-sensitive Urdbean genotypes in growth chamber

Significant genetic variability for the various evaluated traits was recorded in twenty-six selected urdbean genotypes under normal and heat stress condition in both years (Supplementary Table S3). Based on the various physiological and yield and yield related parameters, the following genotypes Mash 114, PantU31,

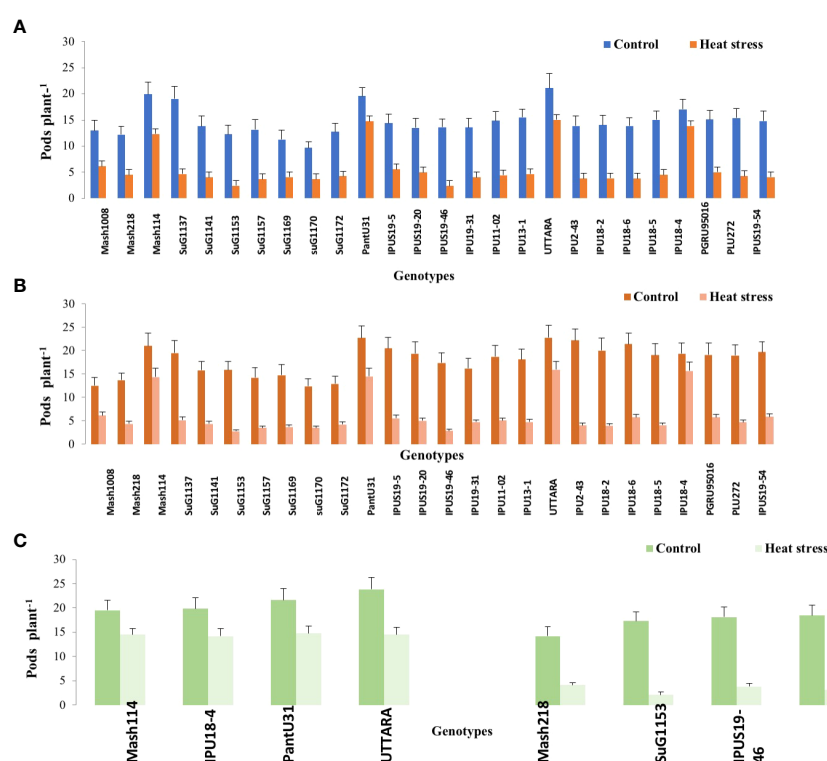


FIGURE 1

Pod number plant⁻¹ of Urdbean genotypes under control (normal-sown; Control) and heat stress environment during 2018 (A), 2019 (B) and in controlled environment of growth chamber (GC; C). LSD values ($P < 0.05$); genotype \times treatment: 2.6 (2018), 3.1 (2019), 3.46 (GC). Values are means \pm SE. ($n = 3$).

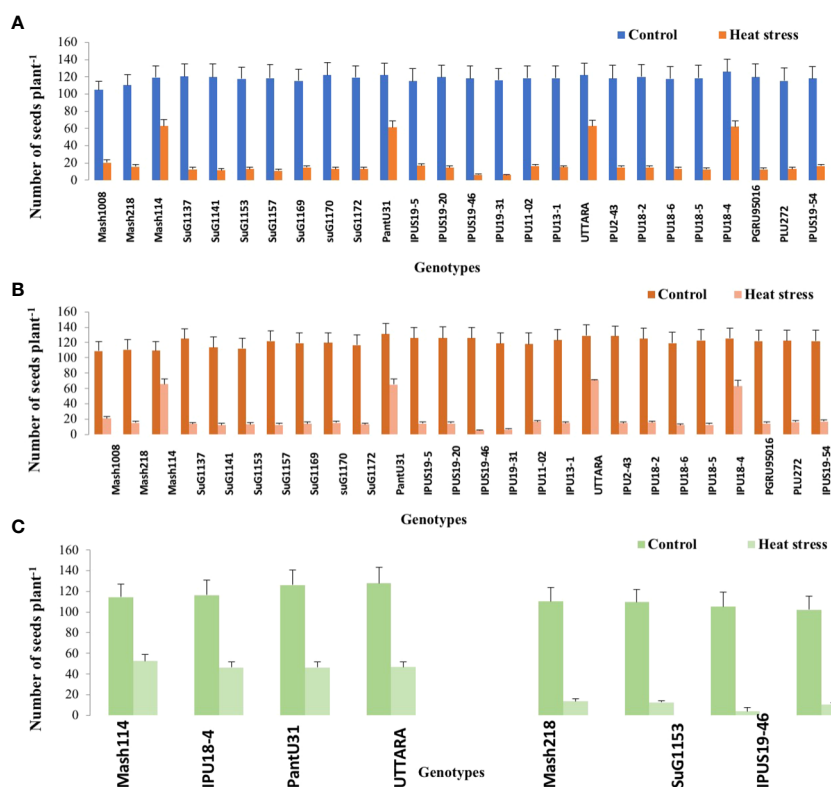


FIGURE 2

Seed number plant⁻¹ of Urdbean genotypes under control (normal-sown; Control) and heat stress environment during 2018 (A), 2019 (B) and in controlled environment of growth chamber (C; GC). LSD values ($P < 0.05$); genotype \times treatment: 6.9 (2018), 7.5 (2019), 12.1 (GC). Values are means \pm SE. ($n = 3$).

IPU18-04 and UTTARA were identified to be heat tolerant for both years under hot environment. Contrastingly, Mash 118, SuG1153, IPU18-5, IPU5(19-46) were identified to be highly sensitive to heat stress for both years.

To validate the response of selected heat tolerant and sensitive genotypes, a selected set of 4 heat tolerant and 4 heat-sensitive urdbean genotypes among the 26 genotypes were examined under growth chamber subjecting them to normal and heat stress treatments, separately. Among these selected 4 heat tolerant genotypes, Mash114 and IPU18-04 revealed high tolerance to heat stress, evidenced by high number of pods plant⁻¹ (14.5, 14.2), high seed number plant⁻¹ (52.5, 46.5) and high efficiency of various physiological traits (chlorophyll content (Supplementary Figure S4), chlorophyll fluorescence (Supplementary Figure S5), stomatal conductance (Supplementary Figure S6), low electrolyte leakage (Supplementary Figure S2), and low malondialdehyde content (Supplementary Figure S7) and reproductive traits [high pollen germination (57.4%, 52.5%) and viability percentage (61.2%, 57.1%)] under heat stress environment (see Supplementary Figure S7).

However, among the heat-sensitive genotypes, IPU5-(19-46) and IPU-18-5 showed high heat stress sensitivity, evidenced by

high reduction of yield and yield-related traits as well as physiological and reproductive traits.

3.4 Correlation analysis

Pod number plant⁻¹ showed significant positive correlations with traits-chlorophyll content, chlorophyll fluorescence, leaf area and stomatal conductance-directly contributing to photosynthesis process whereas significant and negative association of electrolyte leakage (EL) percentage was noticed (Table 2). EL also showed negative correlation with traits (chlorophyll content, chlorophyll fluorescence, leaf area and stomatal conductance), yield traits such as pods plant⁻¹, seeds plant⁻¹, single seed weight and seed yield plant⁻¹ under heat stress environment during both the years.

In urdbean plants grown under growth chamber condition, subjected to heat stress, electrolyte leakage and malondialdehyde (an indicator of oxidative stress) showed highly significant negative correlation with all the physiological traits viz., chlorophyll content, chlorophyll fluorescence, stomatal conductance, relative water content, leaf area, pollen

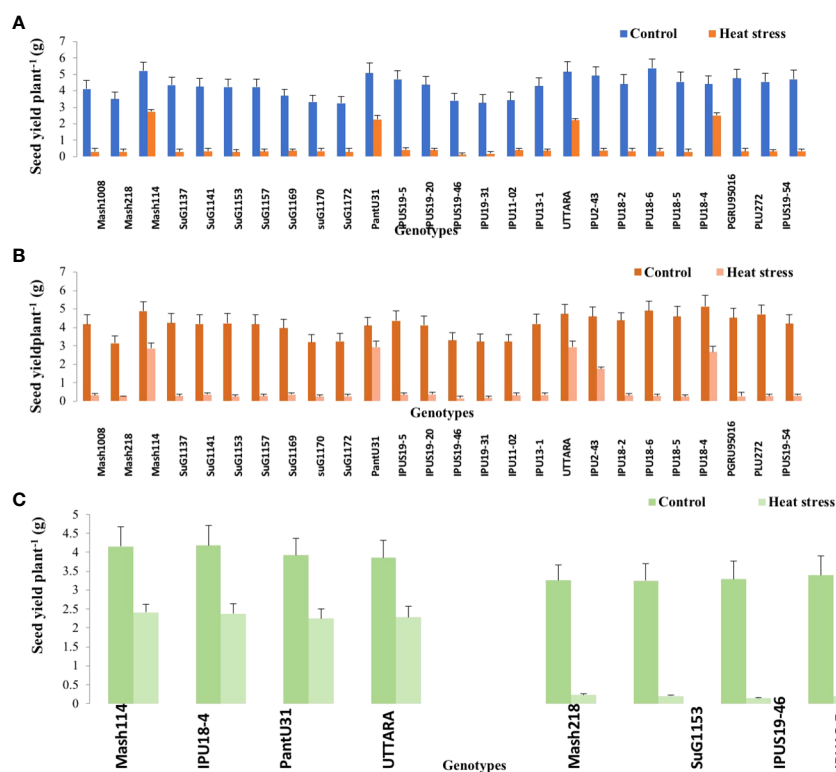


FIGURE 3

Seed yield plant⁻¹ of Urdbean genotypes under control (normal-sown; Control) and heat stress environment during 2018 (A), 2019 (B) and in controlled environment of growth chamber (C; GC). LSD values ($P < 0.05$); genotype \times treatment: 1.3 (2018), 1.5 (2019), 1.3 (GC). Values are means \pm SE. ($n = 3$).

germination % and pollen viability % and yield traits viz. pods plant⁻¹, seed yield plant⁻¹ (Table 2).

High and significant positive correlation of pollen germination % and pollen viability % were noticed with all the traits except electrolyte leakage and malondialdehyde. Likewise, stomatal conductance and RLWC also exhibited high and positive correlation with all the traits except electrolyte leakage and malondialdehyde. The yield and yield related traits viz., pods plant⁻¹, seeds plant⁻¹, single seed weight and seed yield plant⁻¹ showed very strong positive correlation with pollen germination and pollen viability (Table 2) suggesting these traits as vital for screening heat tolerant urdbean genotypes.

3.5 PCA analysis

During the first year, under heat stress environment, PCA analysis (Figure 5) revealed five principal components correlated to 9 traits accounted for 96.5% of total variability. The individual contribution of each component was 76.8%, 6.84%, 5.18%, 4.23% and 3.38%. Analysis of factor loadings of the traits in the retained PCs suggested that seed yield plant⁻¹ (SPY) (13.76), seeds plant⁻¹

(SPP) (13.45) and pods plant⁻¹; PPP (13.12) contributed most positively. In PC2, leaf area (LA) contributed most positively. The trait chlorophyll fluorescence (ChlF) (58.47), chlorophyll (Chl) (76.45) and electrolyte leakage (EL%) (45.1) had highest contribution to PC3, PC4 and PC5, respectively.

Likewise, during second year, PCA analysis (Figure 6) indicated five principal components correlating to 9 traits contributed 97.9% to the total variability. The individual contribution of each component was 81.2%, 9.3%, 3.62%, 2.51% and 1.29%. PPP (12.6) had the highest contribution to PC1. Likewise, LA (50.69) contributed with highest positive value to PC2. EL% (57.38), ChlF (40.18) and Chl (53.1) had highest positive contribution to PC3, PC4 and PC5, respectively.

Under growth chamber, PCA analysis (Figure 7) suggested five PCAs attributing to 13 traits contributing 99.8% to the total variability. The individual contribution of each component was PC1 (97.5%), PC2 (1.03%), PC3 (0.64%), PC4 (0.35%) and PC5 (0.30%). Chlorophyll content (7.81) had the highest contribution to PC1, while single seed weight (32.58) had the highest contribution to PC2. Electrolyte leakage (30.42%) had the highest contribution to PC3 and stomatal conductance (42.09) had the highest contribution to PC4. Seeds plant⁻¹ (32.98) had the highest contribution to PC5.

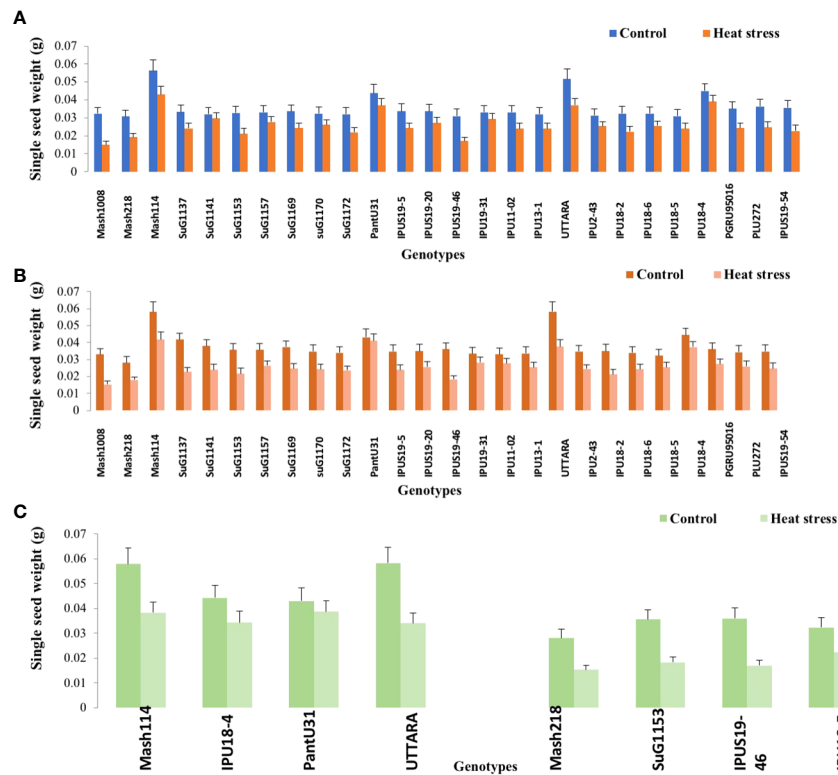


FIGURE 4

Single seed weight of Urdbean genotypes under control (normal-sown; Control) and heat stress environment during 2018 (A), 2019 (B) and in controlled environment of growth chamber (C;GC). LSD values ($P < 0.05$); genotype \times treatment: 0.0023 (2018), 0.0021 (2019), 0.0018 (GC). Values are means \pm SE. ($n = 3$).

TABLE 2 Correlation coefficients of various traits with yield traits in plants under heat stress environment.

Trait	Outdoor environment (2018)		Outdoor environment (2019)		Growth Chamber	
	Number of Pods Plant ⁻¹	Seed yield plant ⁻¹	Number of Pods Plant ⁻¹	Seed yield plant ⁻¹	Number of Pods Plant ⁻¹	Seed yield plant ⁻¹
Electrolyte leakage %	-0.71**	-0.76**	-0.73**	-0.73**	-0.96**	-0.98**
Chlorophyll	0.71**	0.72**	0.80**	0.73**	0.98**	0.99**
Chlorophyll fluorescence	0.68**	0.70**	0.87**	0.82**	0.99**	0.99**
Stomatal Conductance	0.91**	0.95**	0.97**	0.95**	0.98**	0.98**
Leaf area	0.70**	0.67**	0.60**	0.54**	0.98**	0.99**
Malondialdehyde					-0.98**	-0.98**
Pollen viability					0.98**	0.97**
Pollen germination					0.96**	0.99**
Number of Pods Plant ⁻¹	1	0.96**	1	0.95**	1	0.99**
Seed yield plant ⁻¹	0.96**	1	0.92**	1	0.99**	1
Seeds plant ⁻¹	0.98**	0.99**	0.98**	0.95**	0.98**	0.99**
Single seed weight	0.81**	0.88**	0.86**	0.85**	0.98**	0.97**

** denotes significant at 1%.

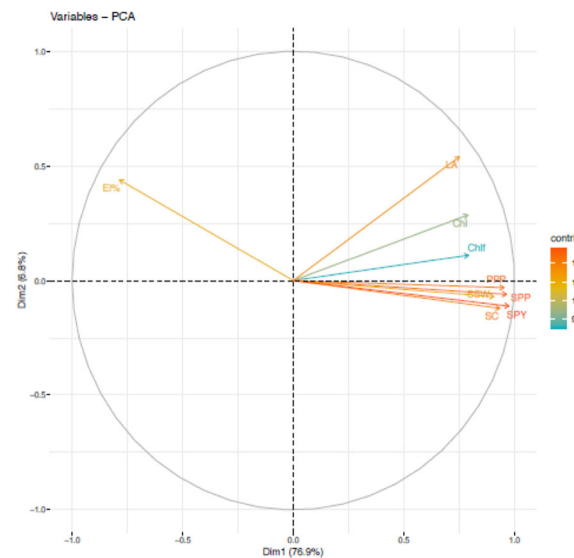


FIGURE 5

Principal component analysis (PCA) of various traits in Urdbean genotypes under heat stress in the year 2018.

3.6 Clustering and identifying heat tolerant Urdbean genotypes based on outdoor experiments

Based on the heat map analysis considering all the physiological and yield-related traits evaluated during the first year in all, the 26 genotypes revealed three major clusters.

Regarding first year, Cluster 1 contained all the highly heat tolerant genotypes, including IPU18-04(43.34%), Mash 114 (47.7%), UTTARA (IPU94-1) (57.39%), and PantU31 (55.45%) (Figure 8) relying on low reduction of seed yield plant⁻¹ (SYP) compared under non-stress and heat stress conditions. The heat-sensitive genotypes viz., Mash218 (91.35% SYP reduction), IPU5 (96.79% SYP reduction),

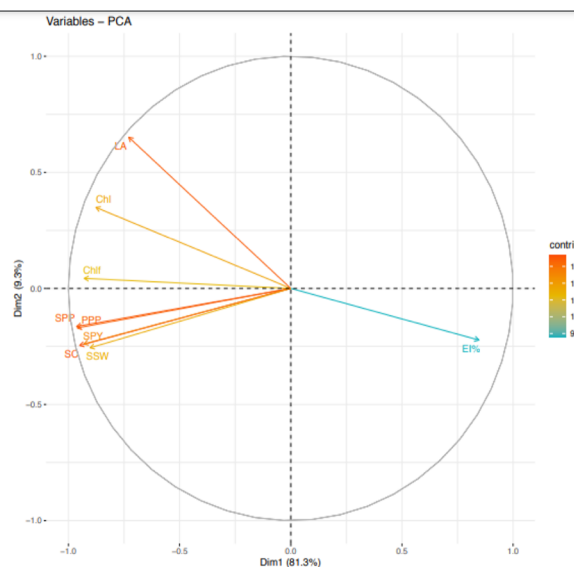


FIGURE 6

Principal component analysis (PCA) of various traits in urdbean genotypes under heat stress in the year 2019.

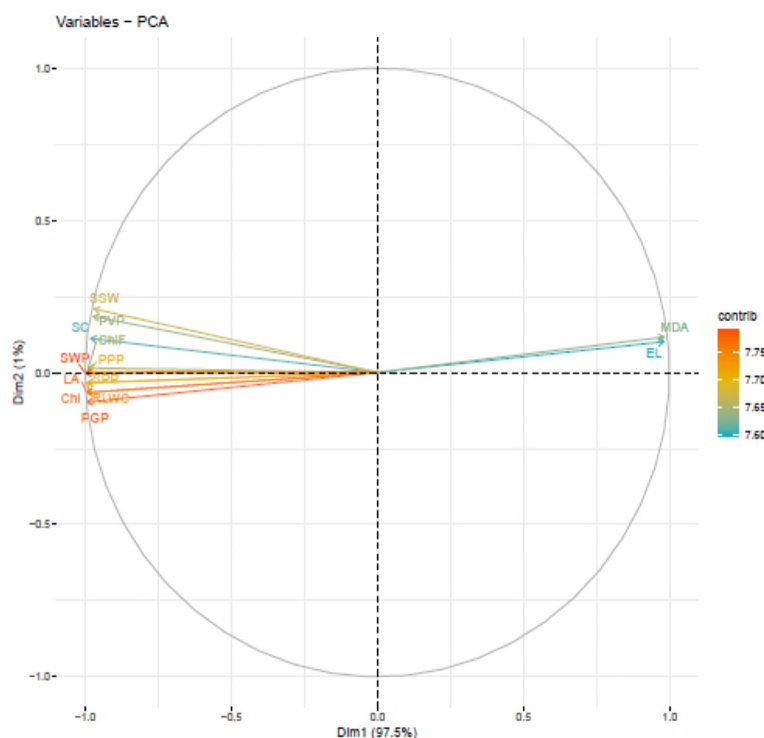


FIGURE 7
Principal component analysis (PCA) of various traits in urbean genotypes under heat stress in a growth chamber.

IPU18-5(93.40% SYP reduction) remained in second cluster. The 3rd cluster had genotypes such as IPU 11-02 (88.45% SYP reduction), SuG1170 (89.79% SYP reduction), SuG1169 (90.235 SYP reduction), Mash 1008 (92.6% SYP reduction)

Likewise, during second year, the heat map also divided the genotypes into three clusters (Figure 9). The first cluster contained the highly heat tolerant genotypes viz., Mash 114 (41.6% SYP reduction), PantU31 (28.7% SYP reduction), IPU18-04 (48.35% SYP reduction) and UTTARA (38% SYP reduction). In the second cluster, all the heat-sensitive genotypes such as Mash 218 (91.5% SYP reduction), SuG1153 (94.1% SYP reduction), IPU18-5 (94.8% SYP reduction) and IPU5 (95.2% SYP reduction) were placed. Other genotypes, for example, IPU2-43 (62.1% SYP reduction), IPU-11-02 (90.3% SYP reduction), and SuG1169 (90.8% SYP reduction) were found in third cluster.

The cluster analysis of the selected genotypes evaluated for various traits under growth chamber condition resulted in two major clusters (Figure 10). The first cluster contained all the four heat-sensitive genotypes such as IPU18-5, IPU5, SuG1153, and Mash 218, whereas the second cluster contained all the heat tolerant genotypes UTTARA, PantU31, IPU18-04, and Mash114.

Various symptoms of heat stress on urbean at vegetative and reproductive growth are shown in Figures 11 and 12.

4 Discussion

Increasing frequency of heat stress events poses serious challenges in all the plant growth stages, especially, reproductive stage, resulting in significant yield loss in various crop plants, including urbean (Jha et al., 2014; Jha et al., 2017; Chaudhary et al., 2020; Chaudhary et al., 2022). Thus, assessing urbean's genetic variability for phenological, morpho-physiological, biochemical and yield and yield related traits is one of the prime objectives for developing heat tolerant climate resilient urbean genotypes.

A selected set of 26 urbean genotype were examined for heat stress tolerance by growing them under non-stress and heat stress under field condition and under controlled growth chamber conditions. Wide range of genetic variability for various physiological, biochemical and yield and yield related traits was observed. Thus, some genotypes showing promising results based on the physiological traits such as chlorophyll content, chlorophyll fluorescence, and yield traits like high pod setting plant⁻¹ and high seed yield plant⁻¹ under heat stress environment were identified and validated for heat tolerance under controlled high temperature environment. These urbean genotypes could be a precious resource for heat tolerance. Further, the tolerant urbean genotypes could be potentially used for investigating the genetic control of heat tolerance in urbean.

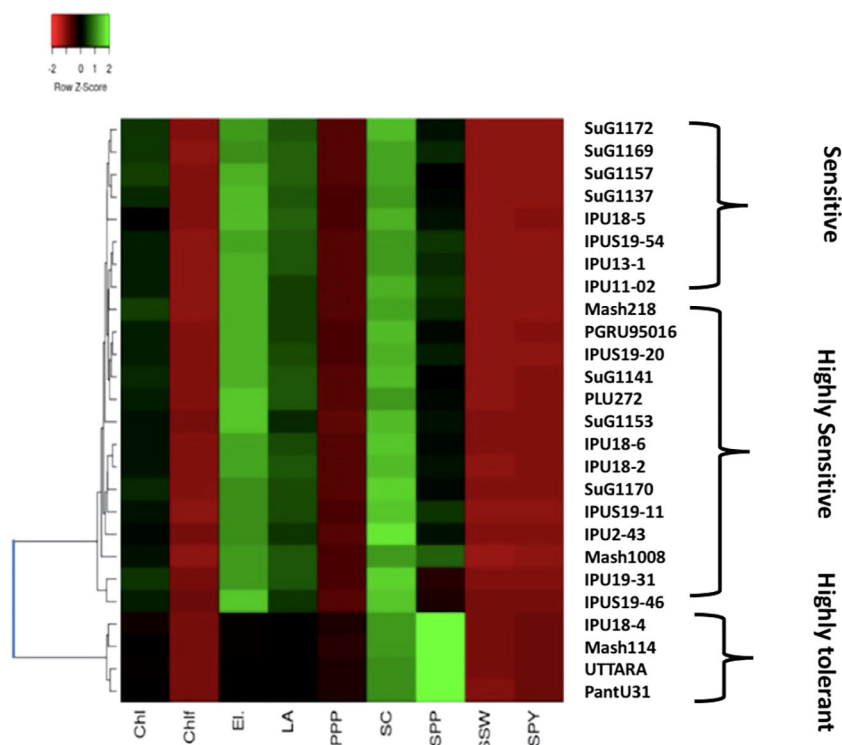


FIGURE 8
Heat map based on the response of urbean genotypes to heat stress in the year 2018.

Among the various physiological traits for assessing heat stress response at vegetative stage during photosynthesis, measuring chlorophyll content, electrolyte leakage, stomatal conductance and photosystem II function are essential parameters for selecting heat tolerant genotypes (Srinivasan et al., 1996; Sita et al., 2017; Bhandari et al., 2020; Devi et al., 2022). Chlorophyll is the main photosynthetic pigment, assists in capturing light energy and enables in photosynthesis process (Wang et al., 2018). Heat stress causes reduction in Chl content resulting in leaf senescence (Kim and Nam, 2007). Improved retention of Chl content under heat stress could be an indication of heat stress tolerance. Thus, genetic variability for Chl content could be useful for selecting heat tolerant genotypes. Heat tolerance based on membrane stability measured by electrolyte/ion leakage is an important trait for selecting genotypes for heat tolerance (Bajji et al., 2002). Under high temperature stress, the genotypes showing low electrolyte leakage indicate stable cell membrane stability and thus are considered heat tolerant. Likewise, stomatal conductance and transpiration cooling are important traits for assessing heat tolerance in plants. Leaf cooling is a vital heat stress avoidance mechanism (Deva et al., 2020) thus, enhanced stomatal conductance and transpiration cooling could help plants to conduct photosynthesis process under high temperature stress

(Porch and Hall, 2013). Thus, genotypes with high stomatal conductance could be heat tolerant. In the current study, Mash114, UTTARA, PantU31 genotypes showed high chlorophyll content, high stomatal conductance, low electrolyte leakage and high Fv/Fm value under heat stress, both under field and growth chamber condition, indicating their heat stress tolerance. Similar findings were reported in chickpea (Devi et al., 2022), lentil (Srinivasan et al., 1996; Delahunty et al., 2015; Sita et al., 2017; Sehgal et al., 2019; Bhandari et al., 2020) and pea (McDonald and Paulsen, 1997) under heat stress.

Of all the growth stages, reproductive stage is the most sensitive stage affected by negative impact of heat stress (Zinn et al., 2010). High temperature stress causes anomalies and malfunction in reproductive processes ranging from reduction in pollen germination (PGP) percentage, pollen viability percentage (PVP), malformation in ovule to inhibition in fertilization process in various crops, including rice (Xu et al., 2021), wheat (Ullah et al., 2022), chickpea (Bhandari et al., 2020; Devi et al., 2022), common bean (Silva et al., 2019; Soltani et al., 2019) and tomato (Gonzalo et al., 2021). High PGP and PVP values are indicators of efficient reproductive function leading to high pod and seed setting resulting in improved yield under heat stress (Firon et al., 2006; Pham et al., 2020). Sufficient range of

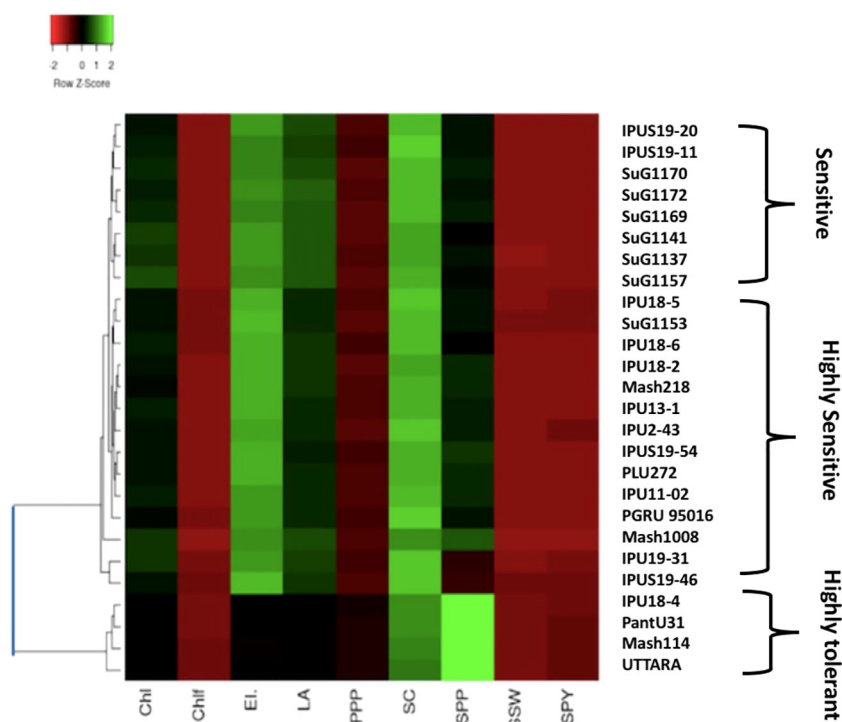


FIGURE 9
Heat map based on the response of urdbean genotypes to heat stress in the year 2019.

genetic variability for PGP (15.4%-57.4%), PVP (24.5%-61.2%) was noted under heat stress in the present study, providing scope for selection and developing heat tolerant urdbean genotypes. Based on these traits, Mash114 and PantU31 genotypes could be promisingly used as donor parents for improving heat tolerance in urdbean. Screening of heat tolerance relying on PGP and PVP has been reported in rice (Zhang et al., 2018), wheat (Bheemanahalli et al., 2019), chickpea (Devi et al., 2022), lentil (Barghi et al., 2013), common bean (Silva et al., 2019; Vargas et al., 2021), tomato (Pham et al., 2020) and sorghum (Djanaguiraman et al., 2018).

Emphasizing on yield and yield-related parameters such as pods plant⁻¹, significant genetic variability ranging from (2.43-15.07 during the first year) and (2.77-15.9 during second year), for seed yield plant⁻¹ (0.11-2.73g during the first year) and (0.16-2.93g during the second year) and for total seeds plant⁻¹ (6.2-62.9 during the first year) and (6.7-71.2 during the second year) under heat stress was recorded. Thus, genotypes with high pod setting, high seed yield plant⁻¹ and high seed number plant⁻¹ under heat stress environment could be promisingly selected as heat tolerant genotypes. Based on these traits, Mash114, PantU31, UTTARA and IPU18-04 were selected as heat tolerant urdbean genotypes. Similarly, in previous studies, based on these yield traits, genotypes “40-10,” “Naparnyk,” and “CDC Meadow” in pea (Jiang et al., 2020), G122, PI

163120, Cornell 503 in common bean (Shonnard and Gepts, 1994; Rainey and Griffiths, 2005a), ICC1205, ICC15614, GNG469, GNG1488, GNG1499, and GNG1969 in chickpea (Devasirvatham et al., 2013), B89-200 and TN88-63 in cowpea (Ehlers and Hall, 1998), 72578, 70548, 71457 and 73838 in lentil (Delahunty et al., 2015) and 55-437 and 796 in groundnut (Ntare et al., 2001) were identified to be heat tolerant.

Correlation studies indicated that electrolyte leakage trait had highly negative association with all the physiological (except MDA) and yield and yield-related traits under heat stress condition, indicating genotypes having high value for electrolyte leakage are highly heat- sensitive genotypes. However, other physiological traits viz., chlorophyll content, chlorophyll fluorescence, relative water content stomatal conductance showing high and positive association with yield and yield related traits viz., seed yield plant⁻¹, total seeds plant⁻¹ and single seed weight indicated that selection of urdbean genotypes with high chlorophyll content, enhanced stomatal conductance and high relative water content under heat stress could be highly heat tolerant. Positive association of chlorophyll content, stomatal conductance trait related to photosynthesis process with yield and yield-related traits ranging from seed yield plant⁻¹, pod number plant⁻¹ and single seed weight under heat stress has been reported in chickpea (Devi et al., 2022), lentil (Sita et al., 2017), and common bean (Petkova et al., 2007).

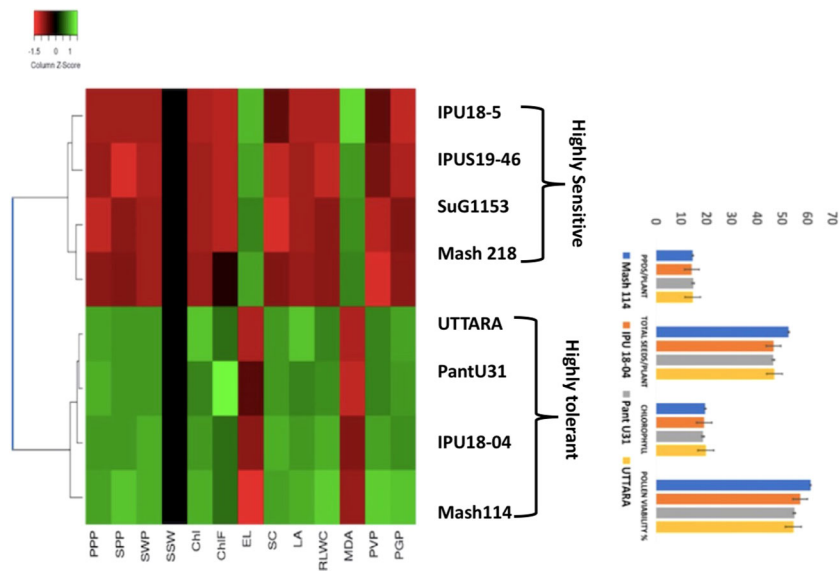


FIGURE 10
Heat map based on the response of urdbean genotypes to heat stress in a growth chamber.



FIGURE 11
Morphological effects of heat stress on Urdbean plants; plant height under control environment (A), reduced plant height under heat stress (HS) environment (B), healthy leaves under control environment (C), scorching of leaves under HS (D), chlorosis in the HS (E), Leaf senescence and abscission in the HS (F).

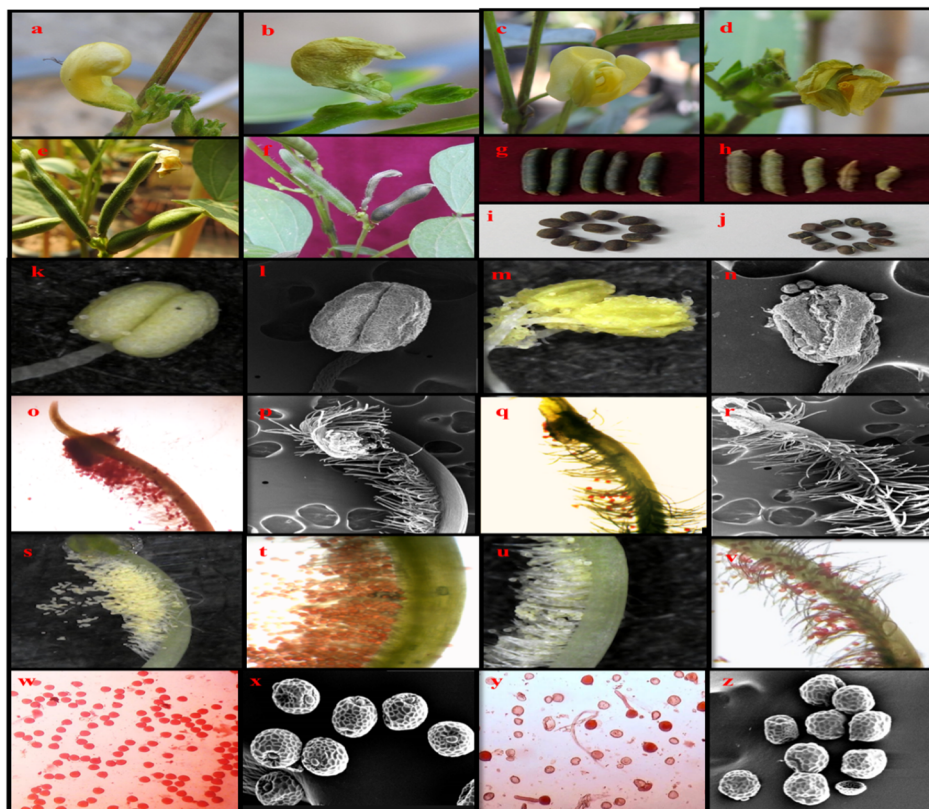


FIGURE 12

Urdbean plants showing various distinctive impacts on it are the reproductive phase when raised under control and heat stress environment. Plants grown under control temperature have healthy bud (A) healthy flower (C) filled pods (egg), normal seeds (I) compact anther (K, L) receptive stigma (O, P) higher pollen load (S, T) and possess viable pollen grains (W, X). However, plants raised under heat stress conditions have more frequency of aborted buds (B) aborted flowers (D) unfilled and aborted pods (F, H) shrivelled seeds (J) distorted anther (M, N) non-receptive stigma (Q, R) less pollen load (U, V) and non-viable pollen grains (Y, Z) healthy flower bud (A) aborted flower bud (B) healthy flower (C) aborted flower (D) healthy pods (E) aborted pod (F) normal pod length (G) reduced pod length (H) healthy seeds (G) shrivelled seeds (H) healthy anther under stereo-microscope (I) healthy anther under SEM (J) distorted anther.

Studies conducted in controlled high temperature environment of growth chamber revealed PGP and PVP to be highly correlated with pods plant⁻¹ and could be used as vital indicators of heat tolerance. Earlier studies have also indicated that these traits could be used for selecting heat tolerant genotypes in chickpea (Devi et al., 2022), common bean (Rainey and Griffiths, 2005), lentil (Sehgal et al., 2019) and tomato (Gonzalo et al., 2021).

High heritability of various morpho-physiological, yield and yield related traits could be of great importance for selecting these traits for screening of heat tolerant genotypes in various crops. High heritability for chlorophyll content, stomatal conductance, seed yield plant⁻¹, pods plant⁻¹, single seed weight, as noticed in the present study has also been noted in heat tolerant chickpea (Jha et al., 2019; Devi et al., 2022), rice (Enzi et al., 2022), tomato (Panthee et al., 2018), wheat (Rebetzke et al., 2013) under high temperature environment.

5 Conclusion

Heat stress related events are becoming serious constraints for crop yield including urdbean thus, causing great concern for global food security. Harnessing the genetic variability for various morpho-physiological and yield and yield related traits existing across the crop gene pool could be one of the important approaches for developing heat tolerant crop cultivars including urdbean. A wide range of genetic variability for various morpho-physiological and yield and yield related traits were captured for a selected 26 urdbean genotypes under both non-stress and heat stress environment for consecutive two years. A selected four heat tolerant and four heat-sensitive genotypes were further validated for their heat stress response under controlled growth chamber condition. Based on the results obtained from both outdoor and controlled growth chamber conditions, yield and yield related traits viz., pods plant⁻¹, seeds plant⁻¹, single seed

weight and seed yield plant⁻¹ showed strong positive correlation with chlorophyll, chlorophyll fluorescence, and stomatal conductance. Similarly, these yield traits had very strong correlation with reproductive traits, pollen germination and pollen viability except electrolyte leakage and malondialdehyde content. These results indicated selection for high pollen germination % and high pollen viability % and yield and yield related traits could assist in selecting heat tolerant urdbean genotypes. Thus, the candidate genotypes PantU31, Mash114, UTTARA and IPU18-04 exhibiting high pod setting and high seed yield plant⁻¹ under heat stress imposed under outdoor and growth chamber environment could be potentially used as heat tolerant donor parents for future urdbean breeding programme. Further, these genotypes can be assessed for their heat tolerance across the multiple locations for confirming their heat tolerance based on various locations.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding authors.

Author contributions

SC conducted the experiment. UJ and PJP helped in analysis and writing part of the manuscript. SS and DS contributed in providing the urdbean lines. KS and SK assisted in conducting the biochemical analyses. HN conceived the idea and helped in writing the manuscript. KHMS helped in writing part of the manuscript and edited the manuscript. All authors contributed to the article and approved the submitted version.

References

- Allakhverdiev, S. I., Kreslavski, V. D., Klimov, V. V., Los, D. A., Carpentier, R., and Mohanty, P. (2008). Heat stress: an overview of molecular responses in photosynthesis. *Photosynth. Res.* 98, 541–550. doi: 10.1007/s1120-008-9331-9330
- Anitha, Y., Vanaja, M., and Kumar, V. G. (2016). Identification of attributes contributing to high temperature tolerance in blackgram (*Vigna mungo* L. hepper) genotypes. *Int. J. Sci. Res.* 5, 1021–1025.
- Awasthi, R., Kaushal, N., Vadez, V., Turner, N. C., Berger, J., Siddique, K. H., et al. (2014). Individual and combined effects of transient drought and heat stress on carbon assimilation and seed filling in chickpea. *Funct. Plant Biol.* 41, 1148–1167. doi: 10.1071/FP13340
- Babicki, S., Arndt, D., Marcu, A., Liang, Y., Grant, J. R., Maciejewski, A., et al. (2016). Heatmapper: web-enabled heat mapping for all. *Nucleic Acids Res.* 44, 147–153. doi: 10.1093/nar/gkw419
- Bajji, M., Kinet, J. M., and Lutts, S. (2002). The use of the electrolyte leakage method for assessing cell membrane stability as a water stress tolerance test in durum wheat. *Plant Growth Regul.* 36, 61–70. doi: 10.1023/A:1014732714549
- Barghi, S. S., Mostafaei, H., Peighami, F., Zakaria, R. A., and Nejhad, R. F. (2013). Response of *in vitro* pollen germination and cell membrane thermostability of lentil genotypes to high temperature. *Intl. J. Agril.* 3 (1), 13.
- 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
- Bhandari, K., Siddique, K. H., Turner, N. C., Kaur, J., Singh, S., Agrawal, S. K., et al. (2016). Heat stress at reproductive stage disrupts leaf carbohydrate metabolism, impairs reproductive function, and severely reduces seed yield in lentil. *J. Crop Improv.* 30, 118–151. doi: 10.1080/15427528.2015.1134744
- Bhandari, K., Sita, K., Sehgal, A., Bhardwaj, A., Gaur, P., Kumar, S., et al. (2020). Differential heat sensitivity of two cool-season legumes, chickpea and lentil, at the reproductive stage, is associated with responses in pollen function, photosynthetic ability and oxidative damage. *J. Agron. Crop Sci.* 206, 734–758. doi: 10.1111/jac.12433
- Bheemanahalli, R., Sunoj, V. J., Saripalli, G., Prasad, P. V., Balyan, H. S., Gupta, P. K., et al. (2019). Quantifying the impact of heat stress on pollen germination, seed set, and grain filling in spring wheat. *Crop Sci.* 59, 684–696. doi: 10.2135/cropsci2018.05.0292
- Bitra, C. E., and Gerats, T. (2013). Plant tolerance to high temperature in a changing environment: scientific fundamentals and production of heat stress-tolerant crops. *Front. Plant Sci.* 4. doi: 10.3389/fpls.2013.00273

Acknowledgments

SC thanks CSIR for project fellowship and CSIR-UGC for providing a doctoral research fellowship during her course study. The corresponding author HN is thankful to DST (FIST and PURSE), UGC, DBT, CSIR, India, The University of Western Australia (Australia), IIPR (Kanpur, India), PAU (Ludhiana, India), for supporting the research work at various times.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1042999/full#supplementary-material>

- Brewbaker, J. L., and Kwack, B. H. (1963). The essential role of calcium ion in pollen germination and pollen tube growth. *Am. J. Bot.* 50, 859–865. doi: 10.1002/j.1537-2197.1963.tb06564.x
- Chaturvedi, P., Wiese, A. J., Ghatak, A., Zaveska-Drabkova, L., Weckwerth, W., and Honys, D. (2021). Heat stress response mechanisms in pollen development. *New Phytol.* 231, 571–585. doi: 10.1111/nph.17380
- Chaudhary, S., Devi, P., Bhardwaj, A., Jha, U. C., Sharma, K. D., Prasad, P. V., et al. (2020). Identification and characterization of contrasting genotypes/cultivars for developing heat tolerance in agricultural crops: Current status and prospects. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.587264
- Chaudhary, S., Devi, P., Hanumantha Rao, B., Jha, U. C., Sharma, K. D., Prasad, P. V., et al. (2022). Physiological and molecular approaches for developing thermotolerance in vegetable crops: a growth, yield and sustenance perspective. *Front. Plant Sci.* 13, 878498. doi: 10.3389/fpls.2022.878498
- Chebolu, 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/s11306-015-0941-941
- Chen, W. L., Yang, W. J., Lo, H. F., and Yeh, D. M. (2014). Physiology, anatomy, and cell membrane thermostability selection of leafy radish (*Raphanus sativus* var. *oleiformis* pers.) with different tolerance under heat stress. *Sci. Hortic.* 179, 367–375. doi: 10.1016/j.scienta.2014.10.003
- Delahunty, A., Nuttall, J., Nicolas, M., and Brand, J. (2015). “Genotypic heat tolerance in lentil,” in *Proceedings of the 17th ASA Conference*. 20–24(Hobart, Australia).
- Devasirvatham, V., Gaur, P., Mallikarjuna, N., Raju, T. N., Trethowan, R. M., and Tan, D. K. Y. (2013). Reproductive biology of chickpea response to heat stress in the field is associated with the performance in controlled environments. *Field Crops Res.* 142, 9–19. doi: 10.1016/j.fcr.2012.11.011
- Deva, C. R., Urban, M. O., Challinor, A. J., Falloon, P., and Svitákova, L. (2020). Enhanced leaf cooling is a pathway to heat tolerance in common bean. *Front. Plant Sci.* 11, 19. doi: 10.3389/fpls.2020.00019
- Devi, P., Jha, U. C., Prakash, V., Kumar, S., Parida, S. K., Paul, P. J., et al. (2022). Response of physiological, reproductive function and yield traits in cultivated chickpea (*Cicer arietinum* L.) under heat stress. *Front. Plant Sci.* 13, 880519. doi: 10.3389/fpls.2022.880519
- Djanaguiraman, M., Perumal, R., Jagadish, S. V. K., Ciampitti, I. A., Welti, R., and Prasad, P. V. V. (2018). Sensitivity of sorghum pollen and pistil to high-temperature stress. *Plant Cell Environ.* 41 (5), 1065–1082. doi: 10.1111/pce.13089
- Ehlers, J. D., and Hall, A. E. (1998). Heat tolerance of contrasting cowpea lines in short and long days. *Field Crops Res.* 55, 11–21. doi: 10.1016/S0378-4290(97)00055-5
- Enzi, V., Ahanchede, W. W., Ayenan, M. A. T., and Ahanchede, A. (2022). Physiological and agronomical evaluation of elite rice varieties for adaptation to heat stress. *BMC Plant Biol.* 22 (1), 1–14. doi: https://doi.org/10.1186/s12870-022-03604-x
- Firon, N., Shaked, R., Peet, M. M., Pharr, D. M., Zamski, E., Rosenfeld, K., et al. (2006). Pollen grains of heat tolerant tomato cultivars retain higher carbohydrate concentration under heat stress conditions. *Sci. Hortic.* 109, 212–217. doi: 10.1016/j.scienta.2006.03.007
- Gonzalo, M. J., Nájera, I., Baixauli, C., Gil, D., Montoro, T., Soriano, V., et al. (2021). Identification of tomato accessions as source of new genes for improving heat tolerance: From controlled experiments to field. *BMC Plant Biol.* 21 (1), 1–28. doi: 10.1186/s12870-021-03104-4
- Heath, R. L., and Packer, L. (1968). Photoperoxidation in isolated chloroplasts: I. kinetics and stoichiometry of fatty acid peroxidation. *Arch. Biochem. Biophys.* 125, 189–198. doi: 10.1016/0003-9861(68)90654-1
- Hedhly, A. (2011). Sensitivity of flowering plant gametophytes to temperature fluctuations. *Environ. Exp. Bot.* 74, 9–16. doi: 10.1016/j.envexpbot.2011.03.016
- Jha, U. C., Bohra, A., Parida, S. K., and Jha, R. (2017). Integrated “omics” approaches to sustain global productivity of major grain legumes under heat stress. *Plant Breed.* 136, 437–459. doi: 10.1111/pbr.12489
- Jha, U. C., Bohra, A., and Singh, N. P. (2014). Heat stress in crop plants: its nature, impacts and integrated breeding strategies to improve heat tolerance. *Plant Breed.* 133, 679–701. doi: 10.1111/pbr.12217
- Jha, U. C., Kole, P. C., and Singh, N. P. (2019). Nature of gene action and combining ability analysis of yield and yield related traits in chickpea (*Cicer arietinum* L.) under heat stress. *Indian J. Agric. Sci.* 89, 500–508.
- Jha, U. C., Nayyar, H., Jha, R., Singh, P. K., Dixit, G. P., Kumar, Y., et al. (2022). “Improving chickpea genetic gain under rising drought and heat stress using breeding approaches and modern technologies,” in *Developing climate resilient grain and forage legumes*. Eds. U. C. Jha, H. Nayyar, S. K. Agrawal and K. H. M. Siddique (Singapore: Springer), 1–25.
- Jiang, Y., Lindsay, D. L., Davis, A. R., Wang, Z., MacLean, D. E., Warkentin, T. D., et al. (2020). Impact of heat stress on pod-based yield components in field pea (*Pisum sativum* L.). *J. Agron. Crop Sci.* 206, 76–89. doi: 10.1111/jac.12365
- Joshi, P. K., and Rao, P. P. (2017). Global pulses scenario: status and outlook. *Ann. N. Y. Acad. Sci.* 1392, 6–17. doi: 10.1111/nyas.13298
- Kaushal, N., Awasthi, R., Gupta, K., Gaur, P., Siddique, K. H., and Nayyar, H. (2013). Heat-stress-induced reproductive failures in chickpea (*Cicer arietinum*) are associated with impaired sucrose metabolism in leaves and anthers. *Funct. Plant Biol.* 40, 1334–1349. doi: 10.1071/FP13082
- Kim, H. J., and Nam, H. G. (2007). Leaf senescence. *Annu. Rev. Plant Biol.* 58, 115–136. doi: 10.1146/annurev.arplant.57.032905.105316
- Li, B., Gao, K., Ren, H., and Tang, W. (2018). Molecular mechanisms governing plant responses to high temperatures. *J. Integr. Plant Biol.* 60, 757–779. doi: 10.1111/jipb.12701
- Lutts, S., Kinet, J. M., and Bouharmont, J. (1996). NaCl-Induced senescence in leaves of rice (*Oryza sativa* L.) cultivars differing in salinity resistance. *Ann. Bot.* 78, 389–398. doi: 10.1006/anbo.1996.0134
- McDonald, G., and Paulsen, G. (1997). High temperature effects on photosynthesis and water relations of grain legumes. *Plant Soil* 196, 47–58. doi: 10.1023/A:1004249200050
- Ntare, B. R., Williams, J. H., and Dougbedji, F. (2001). Evaluation of groundnut genotypes for heat tolerance under field conditions in a sahelian environment using a simple physiological model for yield. *J. Agric. Sci. (Cambridge)* 136, 81–88. doi: 10.1017/S0021859600008583
- Panthee, D. R., Kressin, J. P., and Piotrowski, A. (2018). Heritability of flower number and fruit set under heat stress in tomato. *Hortic. Sci.* 53 (9), 1294–1299. doi: 10.21273/HORTSCI13317-18
- Patterson, H. D., and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554. doi: 10.2307/2334389
- Petkova, V., Denev, I. D., Cholakov, D., and Porjazov, I. (2007). Field screening for heat tolerant common bean cultivars (*Phaseolus vulgaris* L.) by measuring of chlorophyll fluorescence induction parameters. *Sci. Hortic.* 111, 101–106. doi: 10.1016/j.scienta.2006.10.005
- Pham, D., Hoshikawa, K., Fujita, S., Fukumoto, S., Hirai, T., Shinozaki, Y., et al. (2020). A tomato heat-tolerant mutant shows improved pollen fertility and fruit-setting under long-term ambient high temperature. *Environ. Exp. Bot.* 178, 104150. doi: 10.1016/j.envexpbot.2020.104150
- Porch, T. G., and Hall, A. E. (2013). “Heat tolerance,” in *Genomics and breeding for climate-resilient crops* (Berlin, Heidelberg: Springer), 167–202.
- Rainey, K. M., and Griffiths, P. D. (2005). Differential response of common bean genotypes to high temperature. *J. Amer. Hortic. Sci.* 130, 18–23. doi: 10.21273/JASHS.130.1.18
- Rebetzke, G. J., Rattey, A. R., Farquhar, G. D., Richards, R. A., and Condon, A. T. G. (2013). Genomic regions for canopy temperature and their genetic association with stomatal conductance and grain yield in wheat. *Funct. Plant Biol.* 40, 14–33. doi: 10.1071/FP12184
- Riaz, M. W., Yang, L., Yousaf, M. I., Sami, A., Mei, X. D., Shah, L., et al. (2021). Effects of heat stress on growth, physiology of plants, yield and grain quality of different spring wheat (*Triticum aestivum* L.) genotypes. *Sustainability* 13, 1–18. doi: 10.3390/su13052972
- Sehgal, A., Sita, K., Bhandari, K., Kumar, S., Kumar, J., Vara Prasad, P. V., et al. (2019). Influence of drought and heat stress, applied independently or in combination during seed development, on qualitative and quantitative aspects of seeds of lentil (*Lens culinaris* medikus) genotypes, differing in drought sensitivity. *Plant Cell Environ.* 42 (1), 198–211. doi: 10.1111/pce.13328
- Sen Gupta, D., Basu, P. S., Soufmanien, J., Kumar, J., Dhanasekar, P., Gupta, S., et al. (2021). Morpho-physiological traits and functional markers based molecular dissection of heat-tolerance in urdbean. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.719381
- Sharma, L., Priya, M., Bindumadhava, H., Nair, R. M., and Nayyar, H. (2016). Influence of high temperature stress on growth, phenology and yield performance of mungbean [*Vigna radiata* (L.) wilczek] under managed growth conditions. *Sci. Hortic.* 213, 379–391. doi: 10.1016/j.scienta.2016.10.03
- Shonnard, G. C., and Gepts, P. (1994). Genetics of heat tolerance during reproductive development in common bean. *Crop Sci.* 34, 1168–1175. doi: 10.2135/cropsci1994.0011183X003400050005x
- Silva, D. A. D., Pinto-Maglio, C. A. F., Oliveira, É.C.D., Reis, R.L.D.M.D., Carbonell, S. A. M., and Chiorato, A. F. (2019). Influence of high temperature on the reproductive biology of dry edible bean (*Phaseolus vulgaris* L.). *Scientia Agricola* 77, 1–9. doi: https://doi.org/10.1590/1678-992X-2018-0233
- Sita, K., Sehgal, A., Kumar, J., Kumar, S., Singh, S., Siddique, K. H., et al. (2017). Identification of high-temperature tolerant lentil (*Lens culinaris* medik.) genotypes through leaf and pollen traits. *Front. Plant Sci.* 8. doi: 10.3389/fpls.2017.00744
- Soltani, A., Weraduwa, S. M., Sharkey, T. D., and Lowry, D. B. (2019). Elevated temperatures cause loss of seed set in common bean (*Phaseolus vulgaris* L.) potentially through the disruption of source-sink relationships. *BMC Genome* 20, 312. doi: 10.1186/s12864-019-5669-2

- Srinivasan, A., Takeda, H., and Senboku, T. (1996). Heat tolerance in food legumes as evaluated by cell membrane thermostability and chlorophyll fluorescence techniques. *Euphytica* 88 (1), 35–45. doi: 10.1007/BF00029263
- Ullah, A., Nadeem, F., Nawaz, A., Siddique, K. H., and Farooq, M. (2022). Heat stress effects on the reproductive physiology and yield of wheat. *J. Agron. Crop Sci.* 208 (1), 1–17. doi: 10.1111/jac.12572
- Vargas, Y., Mayor-Duran, V. M., Buendia, H. F., Ruiz-Guzman, H., and Raatz, B. (2021). Physiological and genetic characterization of heat stress effects in a common bean RIL population. *PLoS One* 16, e0249859. doi: 10.1371/journal.pone.0249859
- Wahid, A., Gelani, S., Ashraf, M., and Foolad, M. R. (2007). Heat tolerance in plants: an overview. *Environ. Exp. Bot.* 61, 199–223. doi: 10.1016/j.envexpbot.2007.05.011
- Wang, Q. L., Chen, J. H., He, N. Y., and Guo, F. Q. (2018). Metabolic reprogramming in chloroplasts under heat stress in plants. *Int. J. Mol. Sci.* 19, 849. doi: 10.3390/ijms19030849
- Xu, Y., Chu, C., and Yao, S. (2021). The impact of high-temperature stress on rice: challenges and solutions. *Crop J.* 9 (5), 963–976. doi: 10.1016/j.cj.2021.02.011
- Zhang, C., Li, G., Chen, T., Feng, B., Fu, W., Yan, J., et al. (2018). Heat stress induces spikelet sterility in rice at anthesis through inhibition of pollen tube elongation interfering with auxin homeostasis in pollinated pistils. *Rice* 11 (1), 1–12. doi: 10.1186/s12284-018-0206-5
- Zinn, K. E., Tunc-Ozdemir, M., and Harper, J. F. (2010). Temperature stress and plant sexual reproduction: uncovering the weakest links. *J. Exp. Bot.* 61 (7), 1959–1968. doi: 10.1093/jxb/erq053



OPEN ACCESS

EDITED BY

Xi Yin Wang,
North China University of Science and
Technology, China

REVIEWED BY

Quanxi Sun,
Shandong Peanut Research Institute
(CAAS), China
Tingting Chen,
South China Agricultural University,
China

*CORRESPONDENCE

Rajeev K. Varshney
rajeev.varshney@murdoch.edu.au
Weijian Zhuang
weijianz@fafu.edu.cn

SPECIALTY SECTION

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

RECEIVED 19 September 2022

ACCEPTED 22 November 2022

PUBLISHED 04 January 2023

CITATION

Zhang C, Xie W, Fu H, Chen Y,
Chen H, Cai T, Yang Q, Zhuang Y,
Zhong X, Chen K, Gao M, Liu F, Wan Y,
Pandey MK, Varshney RK and
Zhuang W (2023) Whole genome
resequencing identifies candidate
genes and allelic diagnostic markers
for resistance to *Ralstonia*
solanacearum infection in cultivated
peanut (*Arachis hypogaea* L.).
Front. Plant Sci. 13:1048168.
doi: 10.3389/fpls.2022.1048168

COPYRIGHT

© 2023 Zhang, Xie, Fu, Chen, Chen, Cai,
Yang, Zhuang, Zhong, Chen, Gao, Liu,
Wan, Pandey, Varshney and Zhuang.
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(s)
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.

Whole genome resequencing identifies candidate genes and allelic diagnostic markers for resistance to *Ralstonia solanacearum* infection in cultivated peanut (*Arachis hypogaea* L.)

Chong Zhang^{1,2,3}, Wenping Xie^{1,2}, Huiwen Fu^{1,2}, Yuting Chen^{1,2},
Hua Chen^{1,2}, Tiecheng Cai^{1,2}, Qiang Yang^{1,2}, Yuhui Zhuang^{1,2},
Xin Zhong^{1,2}, Kun Chen^{1,2}, Meijia Gao^{1,2}, Fengzhen Liu³,
Yongshan Wan³, Manish K. Pandey⁴,
Rajeev K. Varshney^{1,5*} and Weijian Zhuang^{1,2*}

¹Key Laboratory of Ministry of Education for Genetics, Breeding and Multiple Utilization of Crops, College of Agriculture, Institute of Oil Crops Research, Research Center for Genetics and Systems Biology of Leguminous Oil Plants, Fujian Agriculture and Forestry University, Fuzhou, China, ²State Key Laboratory of Ecological Pest Control for Fujian and Taiwan Crops, Fujian Agriculture and Forestry University, Fuzhou, China, ³State Key Laboratory of Crop Biology, Shandong Key Laboratory of Crop Biology, College of Agronomy, Shandong Agricultural University, Tai'an, China, ⁴Center of Excellence in Genomics and Systems Biology (CEGSB), International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, ⁵Murdoch's Centre for Crop and Food Innovation, State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Murdoch, WA, Australia

Bacterial wilt disease (BWD), caused by *Ralstonia solanacearum* is a major challenge for peanut production in China and significantly affects global peanut field productivity. It is imperative to identify genetic loci and putative genes controlling resistance to *R. solanacearum* (RRS). Therefore, a sequencing-based trait mapping approach termed "QTL-seq" was applied to a recombination inbred line population of 581 individuals from the cross of Yueyou 92 (resistant) and Xihuixiaoli (susceptible). A total of 381,642 homozygous single nucleotide polymorphisms (SNPs) and 98,918 InDels were identified through whole genome resequencing of resistant and susceptible parents for RRS. Using QTL-seq analysis, a candidate genomic region comprising of 7.2 Mb (1.8–9.0 Mb) was identified on chromosome 12 which was found to be significantly associated with RRS based on combined Euclidean Distance (ED) and SNP-index methods. This candidate genomic region had 180 nonsynonymous SNPs and 14 InDels that affected 75 and 11 putative candidate genes, respectively. Finally, eight nucleotide binding site leucine rich repeat (NBS-LRR) putative resistant genes were identified as the important candidate genes with high confidence. Two diagnostic SNP markers

were validated and revealed high phenotypic variation in the different resistant and susceptible RIL lines. These findings advocate the expediency of the QTL-seq approach for precise and rapid identification of candidate genomic regions, and the development of diagnostic markers that are applicable in breeding disease-resistant peanut varieties.

KEYWORDS

peanut, resistance to *Ralstonia solanacearum*, QTL-seq analysis, candidate genes, diagnostic markers

1 Introduction

Bacterial wilt that is caused by *Ralstonia solanacearum* (*R. solanacearum*), is the most damaging bacterial disease that globally affects over 50 and 450 botanical families and plant species, respectively, including several economically important crops such as tobacco, peanut, tomato, and pepper (Salanoubat et al., 2002; Zhang et al., 2017). *R. solanacearum* is a free-living saprophyte that endures in soil and aquatic habitats for long durations (Genin and Boucher, 2004). *R. solanacearum* mostly infects plant roots, propagates in the xylem, disseminates into the stem, and then to the entire plant resulting in wilt and eventual death (Schell, 2000). Bacterial wilt disease often significantly reduces, by 10–30%, the yield and quality of peanut and other important crops; it may also result in complete yield loss (Zhang et al., 2017). Currently, no effective pesticide and biological control method exists to control this pathogen because of its wide host range and durable survival ability (Yu et al., 2011). Nevertheless, cultivating crop varieties that are genetically resistant has efficiently controlled this disease (Sunkara et al., 2014; Reddy, 2016), leading to the development and release of many resistant varieties of the peanut. However, there exists a looming threat of a breakdown of this genetic resistance in China, due to similar resistance mechanisms in both the cultivated varieties (such as Xiekangqing, Taishan Zhenzhu) and wild species (Janila et al., 2016; Luo et al., 2020). Despite the current search for varieties whose resistance is conferred *via* alternate mechanisms, it is imperative to determine the genomic regions and genes that encode resistance to augment the development of new varieties *via* genomics-assisted breeding (GAB) (Pandey et al., 2020).

Through analysis of various plant genomes, map-based cloning of plant genes that confer resistance to *R. solanacearum* was conducted for a few crop species. In *Arabidopsis*, a recessive *RRS1-R* encoding a Tir-NBS-LRR resistant protein with a WRKY domain in resistant line Nd-1 was first identified and cloned by fine mapping (Deslandes et al., 2003). It conferred resistance to GMI1000 when transferred into

Col-5 variants with the dominant susceptible allele. Furthermore, another resistant gene *RPS4* was in the reverse orientation and directly upstream of *RRS1-R*. This physical association triggered host resistance to the pathogen (Narusaka et al., 2009). Relatedly, *RRS1-R* associated with *RPS4* is a dimer that recognizes PopP2 of *R. solanacearum* to trigger RRS (Narusaka et al., 2014). In *Arabidopsis*, three quantitative trait loci (QTLs) for RRS were identified in 100 F₂ recombinant inbred lines (RILs) from another cross of Col-0 × Ler (Godiard et al., 2003). A putative leucine-rich repeat receptor-like kinase (LRR-RLK) gene named *ERECTA* was cloned and found to trigger RRS (Godiard et al., 2003). Recently, in peanut, two genes *AhRRS5* and *AhRLK1* (also known as *AhCLAVATA1*), encoding an NBS-LRR resistance protein and a receptor-like protein kinase, respectively, were identified by reverse genetics. Transgenic tobaccos that overexpressed these two genes conferred a significantly increased level of resistance to RRS, indicating that both *R* genes and RLKs are involved in resistance mechanisms against BWD (Zhang et al., 2017; Zhang et al., 2019). Hitherto, no resistance genes from other plants have been cloned and characterized by the map-based method.

Recently, several QTLs associated with RRS were effectively identified by QTL mapping in many crop species, including tomato (Thoquet et al., 1996; Carmeille et al., 2006; Wang et al., 2013; Shin et al., 2020), pepper (Mimura et al., 2009) (Du et al., 2019), potato (Habe et al., 2019), eggplant (Lebeau et al., 2013; Salgon et al., 2017), tobacco (Wang et al., 2013) and *Medicago truncatula* (Ben et al., 2013). Up to now, both sequencing-based trait mapping and gene discovery techniques are highly utilized due to low sequencing costs and the development of new methods that elucidate genomic loci and candidate genes associated with specific traits (Varshney et al., 2019; Pandey et al., 2020). Such efforts facilitate faster development of diagnostic markers which can be employed in GAB to accelerate the development of new peanut varieties (Pandey et al., 2020). In peanut, Zhao et al. (Zhao et al., 2016) first reported mapping QTL for RRS on the B02 chromosome using a

moderately dense linkage map of 237 SSR and SNP markers. By combining restriction-site-associated DNA sequencing (RAD-seq) and bulk segregant analysis (BSA) techniques, they developed resistant-related SNP markers from the RIL population of crosses between resistant (Yueyou 92) and susceptible (Xinhuixiaoli) varieties. The two detected QTLs (*qBW-1* and *qBW-2*) in the aforementioned RRS study accounted for 21% and 12% of the resistance phenotypic variance in the F_2 generation, respectively. Only two side-by-side QTLs were found at the *qBW-1* locus on the B02 chromosome in the F_8 generation. The resistant resource of Yueyou 92 was from a Chinese landrace Xiekangqing, which is a major source of parental types used for breeding BWD-resistant variants in South China (Janila et al., 2016; Luo et al., 2020).

The rapid QTL-seq approach is critical for identifying genomic regions of a trait of interest in plants and identifies QTLs based on BSA and next-generation sequencing (Takagi et al., 2013). QTL-seq was the preferred choice of a fast and effective method that identifies and maps QTLs of target traits in crop plants (Takagi et al., 2013). For example, it was to identify QTLs of the target trait in rice (Arikrit et al., 2019; Bommisetty et al., 2020; Lei et al., 2020; Yang et al., 2021), cucumber (Lu et al., 2014; Cao et al., 2021; Zhang et al., 2021), chickpea (Das et al., 2014; Singh et al., 2016; Srivastava et al., 2017), tomato (Illa et al., 2015; Topcu et al., 2021), oilseed rape (Wang et al., 2016; Tudor et al., 2020; Dong et al., 2021), maize (Chen et al., 2018; Wang et al., 2021), and peanut (Pandey et al., 2017; Clevenger et al., 2018; Luo et al., 2019; Kumar et al., 2020; Luo et al., 2020; Topcu et al., 2021). In peanut, it was used to map genomic loci and candidate genes for the development of diagnostic markers for RRS in 195 RILs obtained by crossing Yuanza 9102 and Xuzhou 68-4 (Luo et al., 2019). A major and stable QTL (*qBWRB02.1*) on chromosome B02 was identified, which was significantly associated with RRS in three environments. Moreover, two SNP sites were confirmed in diverse breeding lines and cultivars. Unlike Yueyou 92, Yuanza 9102 was derived from the wild species *Arachis diogeni* that was resistant to BWD (Janila et al., 2016; Luo et al., 2019). A stable QTL for RRS was finely mapped via both linkage mapping and QTL-seq tools in a resistant peanut cultivar (Luo et al., 2020). Two hundred and sixty-eight RILs were sequenced, and the phenotypes of variants from the cross between Xuhua 13 (susceptible) and Zhonghua 6 (resistant) among five environments were evaluated. Using both SSR- and SNP-based genetic maps, the QTL *qBWRB02-1* was identified on chromosome B02 as previously reported (Zhao et al., 2016), and this accounted for 37.79–78.86% phenotypic variation across the five environments. Two adjacent candidate QTL regions in the *qBWRB02-1* locus were segmented into *qBWRB02-1-1* (2.81–4.24 Mb) and *qBWRB02-1-2* (6.54–8.75 Mb) (Luo et al., 2020). *qBWRB02-1-1* accounted for 49.43–68.86% phenotypic variation explained (PVE), which was higher than that for *qBWRB02-1-2* (3.96–6.48% PVE). Moreover, this was validated by competitive allele-specific PCR (KASP)

markers in different RILs and natural populations (Luo et al., 2020).

In this study, we utilized a QTL-seq approach to identify concomitant genomic regions, candidate resistance genes and diagnostic markers in a bacterial wilt-resistant peanut variety, Yueyou92. A 7.2 Mb candidate genomic region was elucidated on chromosome 12 significantly associated with RRS. This study reports successful discovery of followed by candidate resistance genes and validated markers for potential use in marker-assisted selection (MAS) for RRS in peanut breeding programs.

2 Materials and methods

2.1 Plant material and growth

Yueyou 92 (YY92), a variety that is highly resistant to BWD, was bred by the Guangdong Academy of Agricultural Sciences, China. It stemmed from Xiekangqing, which was resistant to *R. solanacearum* strains from different parts of China. In comparison, Xinhuixiaoli (XHXL) was a Chinese landrace that was highly susceptible to BWD. Their resistance validation was stable during multiple years of field assessment (Figure 1). A RIL population containing 581 lines was developed from the cross Yueyou 92 × Xinhuixiaoli using the single seed descent (SSD) method. A total of 581 F_{13} RILs were used for trait mapping for RRS. To assess the diagnostic markers, we utilized 18 resistant and 18 susceptible RILs for genotyping using allele-specific markers. All the RILs and parents were cultivated in a field in Yangzhong County (Sanming, Fujian, China).

2.2 Pathogen inoculation and resistance phenotyping

The 581 RILs were evaluated for RRS in three independent crop seasons i.e., in 2016 spring and autumn (2016S and 2016A) and in 2017 spring (2017S). The RILs of F_{11} , F_{12} , and F_{13} generations and parents were cultivated in two-row plots with 20 seeds in each season. One-month to 40-day-old RILs seedlings were inoculated via a previously described artificial method (Zhao et al., 2016; Zhang et al., 2017). Resistance phenotyping of the RILs was performed 25 days after inoculation in the different seasons. Disease symptoms were classified into six disease severity ratings (Figure 1C): (0) = the inoculated leaflets either remained green or were yellow at inoculating sites, but the entire plant was intact and lacked wilt; (1) = the inoculated leaflets had either wilted or fallen off, but the entire plant was intact and lacked wilt; (2) = the main stem/branches of the inoculated leaves had wilt and chlorosis; (3) = the leaves of non-inoculated branches had wilt or were faded green, but the main stem was green; (4) = the entire plant had wilted and died, and all its branches were greenish; and (5) =

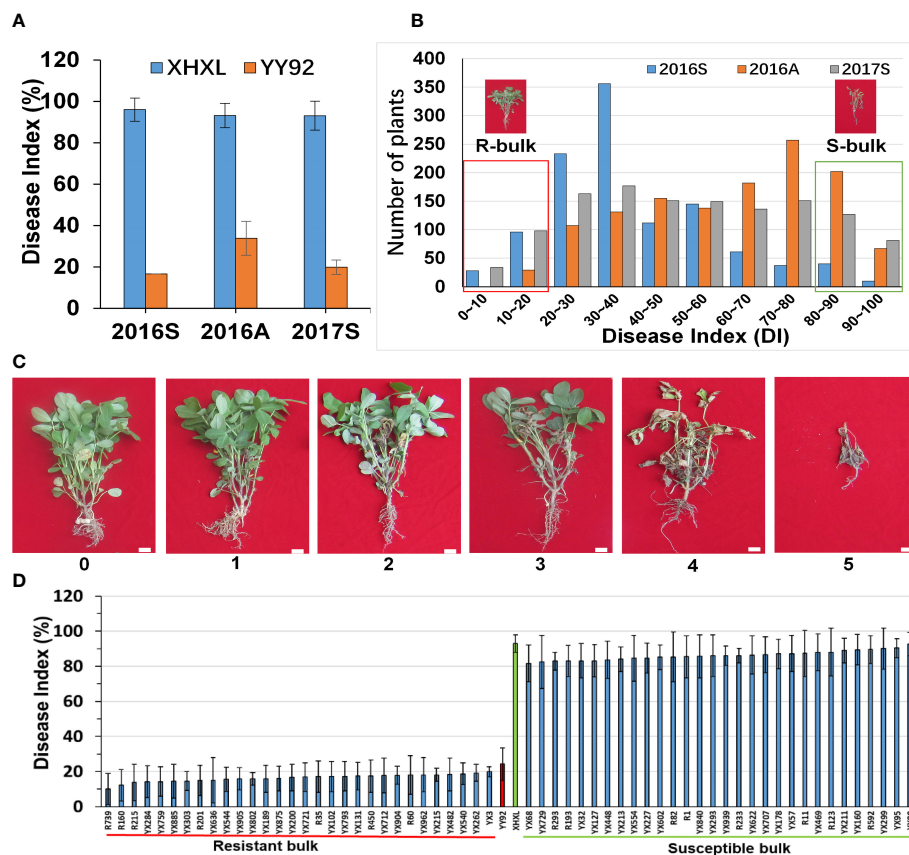


FIGURE 1

Phenotypic variations and construction of the extreme bulks for resistance to *Ralstonia solanacearum* infections. (A) Comparative evaluation of the stability of the resistance between the two parents in the three different crop seasons (2016_Spring, 2016_Autumn, and 2017_Spring). (B) Frequency distribution of disease indexes in the RIL population at three different times. The y-axis represented the number of plants, whereas the x-axis represented the disease index. The red dashed box represented the resistant bulk (R-bulk), and the green dashed box represented the susceptible bulk (S-bulk). (C) Classification of bacterial wilt disease severity. Disease severity was classified using the following scale: 1 = the inoculated leaflets either had wilt or were absent but the entire plant was intact and lacked wilt; 2 = the main stem/branches of the inoculated leaves had wilt and chlorosis; 3 = the lateral branches of the inoculated leaves had wilt or were faded green, but the main stem was green; 4 = the entire plant had wilted and died, but all its branches were green. 5 = the entire plant had wilted and dried up. Bar: 1 cm. (D) Phenotypic variations among the RILs selected for the development of extreme bulks for bacterial wilt resistance. Based on mean values from three environments each with three replications, the 30 RILs with the lowest disease index and the 30 RILs with the highest disease index were used to construct susceptible and resistant bulks.

the entire plant had wilted, dried, and was brownish. The disease index (DI) was calculated using the following formula:

$$\text{Disease index} = \frac{\sum_{i=0}^5 x_i y_i}{x_{\max} \sum y_i} \times 100\%$$

Where, x_i : disease grade value, x_{\max} : the highest disease grade value, and y_i : the number of diseased plants corresponding to the disease rating.

The average DI was calculated for the three replications in a single environment. Statistical analysis of variance (ANOVA) was performed using the DPS7.5 software (Data Processing System, Science Press, China). Values are expressed as the mean \pm standard deviation or standard error as indicated.

Differences between groups were evaluated using one-way ANOVA. Statistical significance was set at $P < 0.05$.

2.3 Extreme bulks construction and whole genome resequencing

The average DI for each RIL was calculated based on phenotyping data from the 2016S, 2016A, and 2017S seasons. We selected 30 resistant and 30 susceptible lines to construct the extreme R/S pool. To develop the resistant bulk (R-Bulk) for RRS, we selected 30 RILs with a low mean disease index and pooled the same amount of DNA from each into one. Similarly,

DNA samples of 30 RILs with a high mean disease index were pooled to construct the susceptible bulk (S-Bulk) for RRS. The genomic DNA of these two extreme pools and those of the two parents was used to construct DNA sequencing libraries. Paired-end reads (151 bp) of four libraries were generated *via* the Illumina HiSeq 2500 platform (Illumina, Inc., USA) with a sequencing depth of approximately 30× of the cultivated peanut genome (~2.7 Gb) for each pool and about 40× for parental plants. The raw sequencing data of the four libraries have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject ID PRJNA851221.

2.4 SNP/InDel genotype detection and annotation

A QTL-seq approach was used to identify the QTLs for RRS (Figure S1) (Takagi et al., 2013). The quality of re-sequenced raw reads from the four libraries was checked. Low-quality reads (those with a proportion of uncalled bases >5%) and adapter sequences were culled. High-quality reads (those with more than 95% nucleotide base calls and high Phred quality scores) were aligned and mapped to the reference genome using the BWA software package (<http://bio-bwa.sourceforge.net/>) with the default parameters (Li and Durbin, 2009). For further analysis, the genome sequences of allotetraploid progenitors of the cultivated peanut *Arachis hypogaea* (Shitouqi) were downloaded from the Peanut Genome Resource website (<http://peanutgr.fafu.edu.cn/>) and used as the reference sequences. Duplicated reads were identified and filtered using Picard (<http://broadinstitute.github.io/picard/>) after mapping the clean reads to the reference genome. To determine the locations and effects of the SNP/InDel variants, we detected and filtered variants in the four libraries using the Genome Analysis Toolkit (GATK, <https://software.broadinstitute.org/gatk/>) (Cingolani et al., 2012) and annotated using SnpEff software (V.5.0e; <https://pcingola.github.io/SnpEff/>) (Reumers et al., 2012).

2.5 Identification of candidate genomic regions

To identify the candidate genomic regions associated with RRS, we further filtered high-quality reads from S-bulk and R-bulk libraries by removing unpaired reads. To equalize the number of reads from each bulk, the filtered reads of the R-bulk were randomly reduced to the same number of the filtered reads of the S-bulk. During the analysis, the SNP sites with genotypes that differed between the two bulks were used to calculate both the sequencing depth of each base in the different bulks and the

Euclidean Distance (ED) value of each site. To eliminate the background noise, the original ED value was processed by power. In this study, the fifth power of the original ED was taken as the correlation value to eliminate the background noise. Then the distance method was used to fit the ED value. For every SNP and InDel in each bulk, ED values were calculated with the formula:

$$ED = \sqrt{(A_{R-bulk} - A_{S-bulk})^2 + (C_{R-bulk} - C_{S-bulk})^2 + (G_{R-bulk} - G_{S-bulk})^2 + (T_{R-bulk} - T_{S-bulk})^2}$$

Each A, G, C, and T letter represented the frequency of its corresponding DNA nucleotide in the resistant and susceptible bulks. The higher the ED value, the stronger the association between the variant with the target characteristic.

The SNP index value was calculated as follows:

$$\text{SNP-index(aa)} = \frac{Maa}{Maa + Paa}$$

$$\text{SNP-index(ab)} = \frac{Mab}{Mab + Pab}$$

$$\Delta\text{SNP-index} = \text{SNP-index(aa)} - \text{SNP-index(ab)}$$

$\Delta\text{SNP-index}$ was calculated by subtracting the SNP-index of the R-bulk from the SNP-index of the S-bulk. SNP-index plots were generated using sliding window analysis with a window size of 2 Mb and increments of 50 Kb. The SNPs with SNP-index <0.3 or read depth <10 in both bulks were culled (Supplementary Figure 1). The SNP index of remaining SNPs as calculated from each bulk was physically plotted onto the 20 cultivated peanut chromosomes. $\Delta\text{SNP-index}$ was calculated by subtracting the SNP index of the resistant bulk from the SNP index of the susceptible bulk. Notably, only those SNPs that had homozygous alleles in both bulks were selected for $\Delta\text{SNP-index}$ calculation. Furthermore, SNP positions were considered as the causal SNPs responsible for the trait of interest if they passed the criterion $\Delta\text{SNP-index} = -1$. $\Delta\text{SNP-index} = -1$ indicated that the allele called in resistant bulk was the same as that of the resistant parent while an alternate base was called in susceptible bulk (Supplementary Figure 3). This analysis was also used in the InDel correlation analysis (Supplementary Figure 4). The DISTANCE method was used to fit $\Delta\text{SNP-indexes}$ and ED values, and the regions above the correlation threshold value (add value) were selected as those related to traits.

The candidate genomic regions related to RRS had the following significant $\Delta\text{SNP/InDel}$ index requirements: $\Delta\text{SNP/InDel}$ index significantly deviated from the statistical confidence intervals under the null hypothesis of no QTLs at a $P < 0.01$ level, and SNP/InDel-index significantly deviated from 0.5 in both bulks. Moreover, the ED values for SNP and InDel were remarkably higher than 0.29 and 0.28, respectively. Finally, the two sets of genomic regions identified from the S-bulk and R-bulk assemblies were combined and considered the genomic regions associated with RRS.

2.6 Diagnostic marker development and validation

To validate the identified genomic regions for RRS, SNPs with different alleles in both bulks and near the intersection terminal were identified and a special marker was developed to narrow the candidate region. For the RIL lines under *R. solanacearum* treatments and with the highest and lowest DI values, we randomly selected 18 of each of these two groups, then extracted DNA from them as well as the parents and the other selected samples. The total volume for the PCR reaction was 20 μ l, comprising DNA template: 50 ng, 2 \times PCR Master Mix: 5 μ l, forward primer: 10 μ M, and reverse primer: 10 μ M. The PCR cycling conditions were as follows: 94°C, 3 min; 30–35 cycles of 94°C, 30 s; 56°C, 30 s; 72°C, 30 s; final extension at 72°C, 10 min. After PCR amplification, the targeted amplicons were identified *via* 1.2% agarose gel electrophoresis.

3 Results

3.1 Phenotype diversity and construction of extreme RRS bulks

To investigate variation in RRS levels of cultivated peanuts, we utilized the resistant “Yueyou 92” (RP) and susceptible “Xinhuixiaoli” (SP) varieties as parents to create multiple generations of segregating RILs populations (Figure 1A). The resistance rate was evaluated based on the severity of *R. solanacearum* infections in RILs, which was calculated as a disease index (DI). The DI value of Yueyou 92 was significantly lower than that of Xinhuixiaoli in three consecutive crop seasons (Figure 1A and Supplementary Table 1). The RILs population had a wide segregation of phenotype variations that formed two peaks of resistance distributions, displaying the main QTLs for RRS regulation (Figure 1B and Supplementary Table 1). Based on the mean values of the disease index in the three field environments, the 30 RILs with the lowest disease index (10.22–20.00%) and the 30 RILs with the highest index (81.68–92.79%) were selected for construction of the resistant (R-bulk) and susceptible bulks (S-bulk) respectively (Figure 1D and Supplementary Table 1). Furthermore, phenotypic identification of R- and S-bulk resistance in the greenhouse was like that in the field environment (Supplementary Figure 1).

3.2 Genome sequencing and SNP/InDel discovery and evaluation

Whole genome sequencing of the parents and the bulks DNA samples was performed on the Illumina HiSeq platform. A total of 114.67 and 103.10 Gb reads were generated for Yueyou

92 and Xinhuixiaoli, and 108.38 and 96.92 Gb for R-bulk and S-bulk respectively. Approximately 97.69% of the reads correctly mapped to the cultivated peanut cv. Shitouqi reference genome (Table 1). An average coverage depth of 42 \times and 37 \times of the reference genome was achieved by Yueyou 92 and Xinhuixiaoli reads respectively, and 38 \times and 34 \times depth for the R-bulk and S-bulk, respectively (Table 1). The mapping results showed that the genome was evenly covered, indicating that the sequencing randomness was good (Supplementary Figure 2).

SNPs/InDels were detected and extracted by the GATK software package. A total of 585,258 SNPs and 167,249 InDels were detected between two parents and 126,900 SNPs and 46,013 InDels were detected between the extreme pools, respectively. The occurrence of the SNPs was 3.5 times more than that of the InDels (Supplementary Table 2; Supplementary Figure 3). After filtering, 381,642 and 98,918 high-quality and homozygous SNP and InDel sites were respectively obtained (Supplementary Table S3). Based on the annotations, 72.7% and 55.4% of the SNPs and InDels, respectively, were in the intergenic region between the extreme pools. Approximately 15% of SNPs and 25% of InDels were upstream and downstream of genes, ~10% of variants in introns, and only ~3% of SNPs and 2.6% of InDels were in the coding region of the two bulks. About 34.6% and 54.7% of the SNPs in the CDS region caused synonymous and nonsynonymous coding variants, respectively. Similarly, Approximately 22.0% of the InDels in the CDS region caused frameshift variants (Supplementary Tables 2, 3).

3.3 Candidate genomic regions for RRS

Using the genome sequences of *Arachis hypogaea* (Shitouqi) as reference, the Euclidean distance (ED) and SNP index, including the Δ SNP-index, were calculated for each genome-wide high-quality SNPs, from RP and SP (Figure 2; Supplementary Figure 6; Supplementary Figure 8). Then, candidate genomic regions for RRS were identified based on ED and Δ SNP-index plots through sliding window analysis of deviations from the threshold value at a 99% confidence level. By using an ED association algorithm, a major peak on Chr12 was identified for RRS, spanning 0–15.19 Mb with an ED > 0.29 ($P < 0.01$) for the SNP. A 6.40 Mb (0.77–7.17 Mb) interval on Chr02 was also identified. By SNP-index and Δ SNP-index, only a genomic interval of 5.83 Mb (4.16–9.99 Mb) on Chr12 deviated from the threshold with the confidence level of $P < 0.01$ (Figure 2; Supplementary Figure 7; Supplementary Figure 9), indicating the interval on Chr12 as the main region controlling the RRS. Moreover, the ED and InDel-indexes (referring to principles of Δ SNP-index) for each identified genomic InDel were calculated for RP and SP. The regions of similarity were confirmed at intervals of 0–7.0 Mb and 0–15 Mb on Chr02 and Chr12, respectively, for ED mapping and 7.49–9.99 Mb on Chr12 for

TABLE 1 Summary of whole genome re-sequencing of parents and bulk lines for bacterial wilt resistance.

Sample ID	Genotype/ bulks	Total_reads	Clean reads	Clean_Base	Q30 (%)	GC (%)	Average depth(X)	Genome coverage ration_1X (%)	Genome coverage ration_5X (%)	Genome coverage ration_10X (%)	Mapped (%)	Properly_mapped (%)
RP	Resistant parent	467,742,004	382,713,001	114,670,878,822	89.89%	36.08%	42	97.72%	97.01%	96.42%	99.55%	96.59%
SP	Susceptible parent	602,985,958	344,100,662	103,101,745,626	90.68%	35.86%	37	97.42%	96.62%	95.77%	99.72%	96.31%
R-bulk	Resistant bulk	602,985,958	361,696,407	108,373,962,086	90.94%	35.99%	38	97.83%	97.11%	96.34%	99.74%	96.23%
S-bulk	Susceptible bulk	630,443,472	323,474,233	96,921,445,922	91.27%	35.94%	34	97.77%	96.99%	95.98%	99.74%	96.84%

The short reads of parents and the extreme bulks were aligned to the genome sequences of cultivated peanut, cv. Shitouqi, Arachis hypogaea Linn (Peanut Genome Resource: <http://peanutgr.fafu.edu.cn/>).

InDel-index association (Supplementary Figures 6-9). Once more, the candidate region on Chr12 was robust with a $P<0.01$ confidence level for both methods. As SNP-indexes enabled fine mapping, the 7.2 Mb (1.8–9.0 Mb) and 5.83 Mb (4.16–9.99 Mb) intervals on Chr12 were collectively identified as candidate region associated with RRS, at 95% and 99% confidence levels, respectively (Figure 3).

3.4 Genetic confirmation of candidate genomic region

To confirm the candidate genomic regions associated with RRS by the QTL-seq approach, we remapped the linkage group (LG) of the existing genetic map to the previously published and newly developed SNP markers. The QTL map had two QTLs located in LG1 (ChrB02) and LG10 in F₂, which explained 21% and 12% of phenotype variations, and one QTL with two adjoining peaks in LG1 (ChrB02) in F₈ (Zhao et al., 2016). The QTL on ChrB02 was located between SNP markers SNP79 and SNP129 in LG1, for which the LOD value was 3.91 and over 6.22, respectively (Zhao et al., 2016). We remapped the SNP markers to the cultivated peanut reference genome. The SNP79 and SNP129 markers were at 1.2 Mb and 9.2 Mb on Chr12, respectively, corroborating identified candidate genomic region through QTL-seq approach (Supplementary Figure 10). Recently, QTL analyses based on SLAF-seq were conducted to detect the candidate QTL region that confers RRS, and the concomitant genotyping and phenotyping data was used for mapping. This resulted in the identification of a consistent region between the SNP marker loci Marker7969064 and Marker7795914 (unpublished data), which explained 45% of the phenotype variations. Marker7969064 and Marke7795914 were located at 6.2 Mb and 8.7 Mb on Chr12, respectively (Supplementary Figure 10). The candidate genomic region associated with RRS as per QTL mapping corroborated that from the QTL-seq method. These results supported QTL-seq results, which revealed the candidate genomic region is associated with RRS.

3.5 Candidate genes associated with RRS

To narrow down the genomic regions and validate effective SNPs associated with RRS on Chr12, we selected a genomic region spanning 7.2 Mb on Chr12 and with 1807 effective SNPs and 629 InDels. Function annotation analysis of the 1807 SNPs revealed 503 intergenic SNPs; 461 intronic; 357 and 225 that were upstream and downstream of genes, respectively; two, six, and eight in 5' UTR, 3' UTR, and splice site regions, respectively; and 67 synonymous and 180 nonsynonymous (two resulted in stop codons) (Supplementary Table 4). Notably, 22 genes with nonsynonymous SNPs were predicted to encode for the NBS-

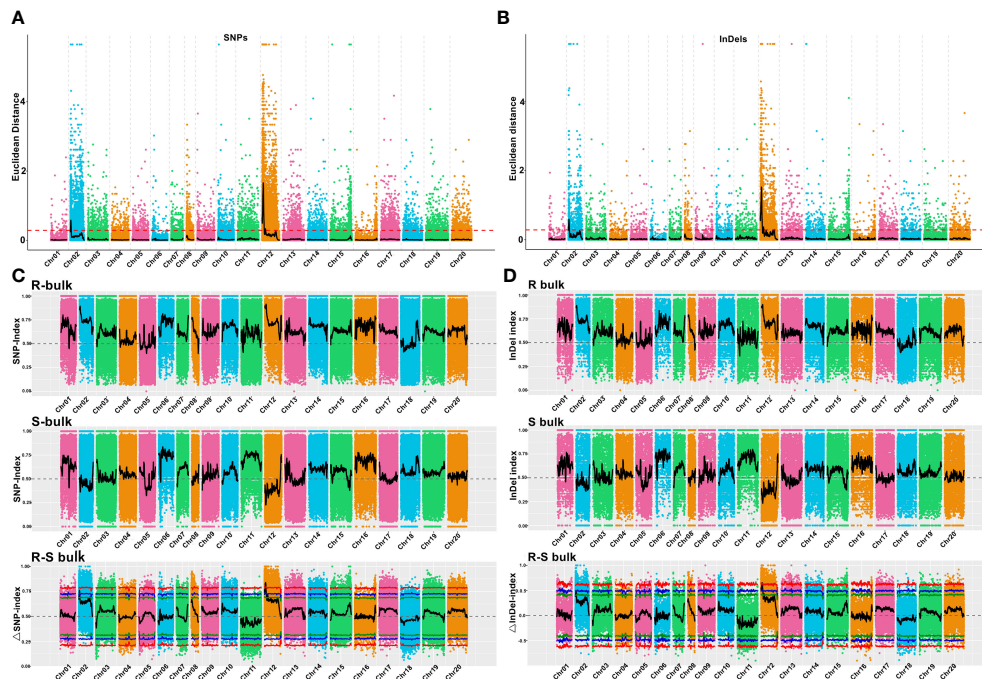


FIGURE 2

Euclidean distance (ED) value distribution of SNPs/InDels and SNP/InDel-index of R- and S-bulks and Δ (SNP/InDel-index) plots generated by sliding-window analyses of 20 cultivated peanut chromosomes. (A, B) Euclidean distance (ED) value distribution of SNPs and InDels in 20 chromosomes. The x-axis represents the chromosome name, the colored dots represent ED values, the black lines represent the fitted ED values (with 2Mb windows sliding in 10 kb steps), and the red dotted lines represent associated thresholds. (C, D) SNP/InDel-index of R- and S-bulks and Δ (SNP/InDel-index) plots generated by sliding-window analysis of cultivated peanut chromosomes. The physical positions of chromosomes are displayed on the X-axis and the average SNP/InDel-index in each 2-Mb physical interval with a 10-kb sliding window is displayed on the Y-axis. Two candidate genomic regions (marked in yellow) were defined using the criteria: average SNP/InDel-index > 0.9 in the R-bulks and average $P < 0.05$. The red, blue, and green lines represent the thresholds at confidence levels of 0.99, 0.95, and 0.90, respectively.

LRR type disease resistance proteins, including *AH12G01510*, *AH12G01540*, *AH12G01550*, *AH12G01560*, *AH12G01570*, *AH12G01600*, *AH12G01900*, *AH12G01920*, *AH12G01980*, *AH12G02020*, *AH12G02090*, *AH12G02120*, *AH12G02130*, *AH12G02310*, *AH12G02330*, *AH12G02370*, *AH12G02390*, *AH12G02410*, *AH12G02880*, *AH12G03230*, *AH12G03600* and *AH12G06320* (Table 2). The *AH12G01460* and *AH12G06300* encode a receptor-like kinase protein and exocyst subunit *Exo70* family protein B2 subunit respectively. *AH12G03290* and *AH12G05320* both encode Serine/threonine-protein phosphatase 7. The other putative candidate genes encoded various kinds of proteins (Table 3). Notably, eight of 22 candidate NBS-LRR resistant genes were identified with high confidence as important candidate genes with the Δ SNP values above 0.60 (Figure 4). Moreover, among the 629 InDels, 152 and 189 were in the intergenic and intronic regions, respectively, 146 and 114 were upstream and downstream of genes, respectively, five and four in 5' UTR and 3' UTR regions, respectively, nine and five resulted in frame shifts and codon changes respectively (Supplementary Table 5). The 180 nonsynonymous SNPs affected 75 putative candidate genes

associated with RRS (Table 2), whereas 14 InDels affected 11 genes (Table 3). Among them, six in NBS-LRR genes *AH12G01920*, *AH12G01980*, *AH12G02090*, *AH12G02390*, *AH12G02440*, *AH12G02600* affected the encoded functions as well as Δ SNP-index results (Table 3). Taken together, these results support the hypothesis that the six NBS-LRR resistance genes might act as the candidate genes related to RRS.

3.6 Allele-specific marker development and validation

To evaluate the specificity of the allele marker of the resistant and susceptible peanut cultivars, we targeted 44 SNPs from candidate NBS-LRR genes for RRS for the development of allele-specific markers (Supplementary Table 6). Allele-specific primers for 44 SNPs were successfully generated. All 44 allele-specific markers were checked for polymorphisms between parental genotypes of the RIL population (YY92 and XHXL). Of the 44 markers, 30 allele markers had good amplification, whereas 14 markers did not yield amplicons with clear bands



FIGURE 3

Genome-wide summary of the putative genomic regions associated with resistance to *Ralstonia solanacearum* infections. (A) Chromosomes of the cultivated peanut reference genome. (B) The genome-wide density of total genes. (C, D) Plots of SNP-index of R- and S-bulks generated by sliding-window analyses of cultivated peanut chromosomes. (E) Δ SNP-index plot generated by using the Shitouqi assembly as a reference genome. Label definitions from outside to inside: upper probability values at 99% (orange) and 95% (green) confidence levels. Δ SNP-index, lower probability values at 95% (green), and 99% (orange) confidence levels. (F, G) InDel-index of R- and S-bulks plots generated by sliding-window analyses of cultivated peanut chromosomes. (H) Δ InDel-index plot generated by using the Shitouqi assembly as a reference genome. Label definitions from outside to inside: upper probability values at 99% (orange) and 95% (green) confidence levels. Δ SNP-index, lower probability values at 95% (green) and 99% (orange) confidence levels.

from samples of parental genotypes. Of the 30 amplified markers, two markers (*qRRS18* and *qRRS19*) in *AH12G03230* and *AH12G06320* genes co-segregated with RRS and may thus be deployed for RRS breeding (Figure 5). These two polymorphic markers were validated on a panel of diverse genotypes containing the resistant parent (Yueyou 92), 18 introgression-resistant RIL lines (YX131, YX189, YX284, YX303, YX636, YX712, YX759, YX905, YX962, YX540, YX544, YX793, YX802, YX875, R160, R201, R215 and R739), the susceptible parent (Xinhuixiaoli), and 18 susceptible RIL lines (YX32, YX68, YX160, YX211, YX293, YX554, YX622, YX840, YX178, R123, R592, YX57, YX80, YX95, YX299, YX469, YX707 and YX939). The primers for the diagnostic marker '*qRRS18*' amplified a 302-bp fragment in the susceptible parent and different susceptible RIL lines, but none in the resistant genotypes (Figure 5A). In contrast, primers for another

diagnostic marker '*qRRS19*' amplified 217-bp fragment in the resistant genotypes and none in susceptible lines (Figure 5G). Most importantly, these two diagnostic markers (*qRRS18* + *qRRS19*) could be further developed and used to distinguish between homozygotes and heterozygotes in the segregating population; i.e., susceptible lines will have a 302-bp allele from the marker '*qRRS18*' and resistant lines will have a 217-bp allele from the marker '*qRRS19*'. These two markers can be used as diagnostic marks for breeding resistant bacterial wilt varieties *via* MAS approach.

4 Discussion

With the advent of complete genome sequencing of diploid progenitor species and cultivated tetraploid variants, the QTL-

TABLE 2 Identification of SNPs in putative candidate genes in the genomic region for resistance to on chromosome 12.

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref, Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref, Alt	SNP-index of SB	Delta SNP-index (RB SNP-index-SB SNP-index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G01450	Chr12	1817271	A	G	R	25	21,4	0.84	R	10	1,9	0.10	0.74	NON_SYNONYMOUS_CODING	Tct/Cct	0.459825	-0.460953	0.585964	-0.588628	AT-rich interactive domain-containing protein 1
AH12G01450	Chr12	1823428	C	T	Y	29	25,4	0.86	Y	15	3,12	0.20	0.66	NON_SYNONYMOUS_CODING	Gaa/Aaa	0.459812	-0.460941	0.585948	-0.588612	AT-rich interactive domain-containing protein 1
AH12G01460	Chr12	1832855	C	T	Y	19	15,4	0.79	Y	10	1,9	0.10	0.69	NON_SYNONYMOUS_CODING	Gaa/Aaa	0.4598	-0.460929	0.585932	-0.588596	Receptor-like protein B6:U6kinase 4
AH12G01510	Chr12	1864236	C	G	C	19	19,0	1.00	S	8	6,2	0.75	0.25	NON_SYNONYMOUS_CODING	Caa/Gaa	0.459761	-0.460891	0.585885	-0.588548	Putative disease resistance RPP13-like protein 1
AH12G01540	Chr12	1888227	G	C	G	17	16,1	0.94	S	5	3,2	0.60	0.34	NON_SYNONYMOUS_CODING	Gga/Cga	0.459735	-0.460866	0.585852	-0.588516	Putative disease resistance RPP13-like protein 1
AH12G01550	Chr12	1889601	G	A	G	19	19,0	1.00	R	5	2,3	0.40	0.60	NON_SYNONYMOUS_CODING	Gaa/Aaa	0.459735	-0.460866	0.585852	-0.588516	Putative disease resistance RPP13-like protein 1
AH12G01550	Chr12	1889607	A	C	A	20	20,0	1.00	M	6	2,4	0.33	0.67	NON_SYNONYMOUS_CODING	Aat/Cat	0.459735	-0.460866	0.585852	-0.588516	Putative disease resistance RPP13-like protein 1
AH12G01550	Chr12	1889737	T	A	T	22	22,0	1.00	W	5	3,2	0.60	0.40	NON_SYNONYMOUS_CODING	gTt/gAt	0.459735	-0.460866	0.585852	-0.588516	Putative disease resistance RPP13-like protein 1
AH12G01560	Chr12	1889759	T	A	T	17	17,0	1.00	W	5	4,1	0.80	0.20	NON_SYNONYMOUS_CODING	gaT/gaA	0.459735	-0.460866	0.585852	-0.588516	Putative disease resistance RPP13-like protein 1
AH12G01570	Chr12	1896280	A	G	R	21	19,2	0.90	R	5	2,3	0.40	0.50	NON_SYNONYMOUS_CODING	gAt/gGt	0.459722	-0.460854	0.585837	-0.5885	Putative disease resistance RPP13-like protein 1
AH12G01570	Chr12	1896290	G	A	R	22	20,2	0.91	R	5	2,3	0.40	0.51	NON_SYNONYMOUS_CODING	atG/atA	0.459722	-0.460854	0.585837	-0.5885	Putative disease resistance RPP13-like protein 1
AH12G01570	Chr12	1896327	G	T	G	17	16,1	0.94	K	4	1,3	0.25	0.69	NON_SYNONYMOUS_CODING	Ggc/Tgc	0.459722	-0.460854	0.585837	-0.5885	Putative disease resistance RPP13-like protein 1
AH12G01570	Chr12	1896685	G	A	G	14	13,1	0.93	R	5	3,2	0.60	0.33	NON_SYNONYMOUS_CODING	cGa/cAa	0.459722	-0.460854	0.585837	-0.5885	Putative disease resistance RPP13-like protein 1
AH12G01570	Chr12	1896707	C	A	C	16	15,1	0.94	M	5	3,2	0.60	0.34	NON_SYNONYMOUS_CODING	ttC/ttA	0.459722	-0.460854	0.585837	-0.5885	Putative disease resistance RPP13-like protein 1
AH12G01600	Chr12	1929770	C	G	C	23	23,0	1.00	S	11	3,8	0.27	0.73	NON_SYNONYMOUS_CODING	Ctt/Gtt	0.459683	-0.460815	0.585788	-0.588452	Putative disease resistance RPP13-like protein 1
AH12G01600	Chr12	1929827	T	A	T	22	22,0	1.00	W	6	2,4	0.33	0.67	NON_SYNONYMOUS_CODING	Tgt/AgT	0.459683	-0.460815	0.585788	-0.588452	Putative disease resistance RPP13-like protein 1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB-SB-SNP-index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12C01600	Chr12	1929839	T	G	T	20	20.0	1.00	K	6	2.4	0.33	0.67	NON_SYNONYMOUS_CODING	Tct/Gct	0.459683	-0.460815	0.585788	-0.588452	Putative disease resistance RPPI3-like protein 1
AH12C01600	Chr12	1931628	A	T	A	15	15.0	1.00	W	5	2.3	0.40	0.60	NON_SYNONYMOUS_CODING	aAg/aTg	0.45967	-0.460802	0.585771	-0.588435	Putative disease resistance RPPI3-like protein 1
AH12C01600	Chr12	1931686	A	C	A	12	12.0	1.00	M	11	3.8	0.27	0.73	NON_SYNONYMOUS_CODING	ttA/tC	0.45967	-0.460802	0.585771	-0.588435	Putative disease resistance RPPI3-like protein 1
AH12C01600	Chr12	1931802	C	T	C	13	13.0	1.00	Y	9	3.6	0.33	0.67	NON_SYNONYMOUS_CODING	tC ₂ g/tTg	0.45967	-0.460802	0.585771	-0.588435	Putative disease resistance RPPI3-like protein 1
AH12C01630	Chr12	1962973	T	C	Y	19	14.5	0.74	Y	14	4.10	0.29	0.45	NON_SYNONYMOUS_CODING	aAa/aCa	0.45963	-0.460764	0.585722	-0.588387	Penatriopptide repeat-containing protein
AH12C01670	Chr12	2019013	A	C	A	24	23.1	0.96	M	21	14.7	0.67	0.29	NON_SYNONYMOUS_CODING	aaA/aaC	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019033	A	C	A	25	24.1	0.96	M	19	13.6	0.68	0.28	NON_SYNONYMOUS_CODING	Aca/Cca	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019086	T	C	Y	22	20.2	0.91	Y	23	16.7	0.70	0.21	NON_SYNONYMOUS_CODING	Ttu/Ctt	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019242	A	G	A	16	15.1	0.94	R	20	18.2	0.90	0.04	NON_SYNONYMOUS_CODING	Aat/Gat	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019615	A	G	A	21	21.0	1.00	R	14	13.1	0.93	0.07	NON_SYNONYMOUS_CODING	aAg/aCg	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019677	A	G	R	27	25.2	0.93	R	16	11.5	0.69	0.24	NON_SYNONYMOUS_CODING	Aat/Gat	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019736	C	G	S	24	22.2	0.92	S	20	13.7	0.65	0.27	NON_SYNONYMOUS_CODING	ttC/tTG	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019773	G	A	G	25	24.1	0.96	R	20	14.6	0.70	0.26	NON_SYNONYMOUS_CODING	Gtt/Att	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2019997	T	G	T	21	21.0	1.00	K	11	6.5	0.55	0.45	NON_SYNONYMOUS_CODING	atT/atG	0.459384	-0.460528	0.58544	-0.588095	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2020027	G	T	G	21	21.0	1.00	K	11	9.2	0.82	0.18	NON_SYNONYMOUS_CODING	agC/agT	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2020059	G	T	G	21	21.0	1.00	K	9	7.2	0.78	0.22	NON_SYNONYMOUS_CODING	aCa/aTa	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12C01670	Chr12	2020110	T	C	T	22	21.1	0.95	Y	22	15.7	0.68	0.27	NON_SYNONYMOUS_CODING	gTa/gCa	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB-SNP-index-SB-SNP-index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	U95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	U99 (99% confidence interval lower side)	Gene function
AH12G01670	Chr12	2020205	G	A	R	27	24.3	0.89	R	31	17.14	0.55	0.34	NON_SYNONYMOUS_CODING	G1a/Ala	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12G01670	Chr12	2020297	G	C	S	30	27.3	0.90	S	28	15.13	0.54	0.36	NON_SYNONYMOUS_CODING	gaG/gaC	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12G01670	Chr12	2020334	A	G	A	33	30.3	0.91	R	30	19.11	0.63	0.28	NON_SYNONYMOUS_CODING	Act/Gat	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12G01670	Chr12	2020367	A	G	A	31	30.1	0.97	R	26	20.6	0.77	0.20	NON_SYNONYMOUS_CODING	Alg/Gtg	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12G01670	Chr12	2020380	G	A	G	26	25.1	0.96	R	26	20.6	0.77	0.19	NON_SYNONYMOUS_CODING	aGg/aAg	0.459292	-0.46044	0.585336	-0.587987	HXXXD-type acyltransferase family protein
AH12G01690	Chr12	2034547	T	C	T	25	24.1	0.96	Y	20	17.3	0.85	0.11	NON_SYNONYMOUS_CODING	Ttu/Ctt	0.459047	-0.460203	0.585059	-0.587699	HXXXD-type acyltransferase family protein
AH12G01720	Chr12	2106320	T	C	Y	17	13.4	0.76	Y	13	2.11	0.15	0.61	NON_SYNONYMOUS_CODING	Agf/Ggt	0.458788	-0.459954	0.584757	-0.58739	Subtilase family protein
AH12G01780	Chr12	2249742	C	T	C	18	16.2	0.89	Y	17	3.14	0.18	0.71	NON_SYNONYMOUS_CODING	aCa/aTa	0.458144	-0.459339	0.584017	-0.586617	Glycerol-3-phosphate 2-O-acyltransferase 6
AH12G01880	Chr12	2427316	C	T	Y	26	23.3	0.88	Y	13	3.10	0.23	0.65	NON_SYNONYMOUS_CODING	ttC/tTc	0.457359	-0.45859	0.583129	-0.585683	UDP-glycosyltransferase 91A1
AH12G01900	Chr12	2432994	A	G	R	23	20.3	0.87	R	15	2.13	0.13	0.74	NON_SYNONYMOUS_CODING	cAc/cCc	0.457315	-0.458548	0.58308	-0.585627	Putative disease resistance RPPI3-like protein 1
AH12G01910	Chr12	2448354	A	G	R	25	22.3	0.88	R	14	1.13	0.07	0.81	NON_SYNONYMOUS_CODING	Agg/Gga	0.457272	-0.458507	0.58303	-0.585572	UDP-glycosyltransferase 91A1
AH12G01920	Chr12	2473114	G	A	R	26	23.3	0.88	R	15	2.13	0.13	0.75	NON_SYNONYMOUS_CODING	gGa/gAa	0.457162	-0.458401	0.582905	-0.585432	Putative disease resistance protein A13g14460
AH12G01920	Chr12	2473124	G	C	S	24	21.3	0.88	S	19	3.16	0.16	0.72	NON_SYNONYMOUS_CODING	tgG/tgC	0.457162	-0.458401	0.582905	-0.585432	Putative disease resistance protein A13g14460
AH12G01980	Chr12	2534393	C	T	Y	19	17.2	0.89	Y	19	17.2	0.89	0.00	NON_SYNONYMOUS_CODING	cCa/cTa	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12G01980	Chr12	2534405	C	T	Y	22	20.2	0.91	Y	18	16.2	0.89	0.02	NON_SYNONYMOUS_CODING	cCi/cTi	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12G01980	Chr12	2534443	A	G	R	16	15.1	0.94	R	17	15.2	0.88	0.06	NON_SYNONYMOUS_CODING	Agg/Gga	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12G01980	Chr12	2534503	C	T	Y	18	16.2	0.89	Y	14	10.4	0.71	0.17	NON_SYNONYMOUS_CODING	Ccc/Tcc	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref, Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref, Alt	SNP-index of SB	Delta SNP-index (RB-SNP-index-SB-SNP-index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12C01980	Chr12	2534516	C	T	Y	19	16.3	0.84	Y	13	8.5	0.62	0.23	NON_SYNONYMOUS_CODING	tCg/tTg	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2534533	C	T	Y	22	18.4	0.82	Y	14	9.5	0.64	0.18	NON_SYNONYMOUS_CODING	CaI/Tca	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2534560	G	A	R	22	18.4	0.82	R	15	10.5	0.67	0.15	NON_SYNONYMOUS_CODING	Gta/Ata	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2539578	T	A	T	24	24.0	1.00	W	23	20.3	0.87	0.13	NON_SYNONYMOUS_CODING	Aca/Tca	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2539652	G	C	G	20	20.0	1.00	S	20	15.5	0.75	0.25	NON_SYNONYMOUS_CODING	Gtc/Ctc	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2539689	C	A	C	26	26.0	1.00	M	24	15.9	0.63	0.38	NON_SYNONYMOUS_CODING	CaI/Aca	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2539742	G	C	G	27	27.0	1.00	S	29	20.9	0.69	0.31	NON_SYNONYMOUS_CODING	tCg/tGg	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2539781	T	A	T	24	24.0	1.00	W	33	22.11	0.67	0.33	NON_SYNONYMOUS_CODING	aAc/aTc	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2539838	A	T	A	21	20.1	0.95	W	27	18.9	0.67	0.29	NON_SYNONYMOUS_CODING	gaT/gaA	0.456951	-0.458197	0.582663	-0.585158	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2540304	C	G	C	31	31.0	1.00	S	16	12.4	0.75	0.25	NON_SYNONYMOUS_CODING	Gat/Cat	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2540329	G	C	G	29	29.0	1.00	S	15	11.4	0.73	0.27	NON_SYNONYMOUS_CODING	aat/aG	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2540382	T	G	T	25	25.0	1.00	K	14	11.3	0.79	0.21	NON_SYNONYMOUS_CODING	Aac/Cac	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2545381	A	C	A	23	22.1	0.96	M	17	14.3	0.82	0.13	NON_SYNONYMOUS_CODING	Ttc/Gcc	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2545481	A	G	R	23	20.3	0.87	R	18	12.6	0.67	0.20	NON_SYNONYMOUS_CODING	gAc/gGc	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C01980	Chr12	2545502	C	G	S	26	23.3	0.88	S	15	12.3	0.80	0.08	NON_SYNONYMOUS_CODING	aCta/Gt	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C02020	Chr12	2547669	C	T	C	13	13.0	1.00	Y	19	15.4	0.79	0.21	NON_SYNONYMOUS_CODING	gCa/gAa	0.45693	-0.458177	0.582639	-0.58513	Putative disease resistance RPPI3-like protein 1
AH12C02090	Chr12	2620498	C	T	C	25	25.0	1.00	Y	9	7.2	0.78	0.22	NON_SYNONYMOUS_CODING	Gac/Aac	0.456824	-0.458079	0.582535	-0.584984	Putative disease resistance RPPI3-like protein 1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB-SNP-index-SB)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G02090	Chr12	2620504	A	C	A	25	25.0	1.00	M	9	7.2	0.78	0.22	NON_SYNONYMOUS_CODING	Tac/Gac	0.456824	-0.458079	0.582535	-0.584984	Putative disease resistance RPPI3-like protein 1
AH12G02090	Chr12	2621513	C	A	M	22	19.3	0.86	M	13	8.5	0.62	0.25	NON_SYNONYMOUS_CODING	aGg/aTg	0.456824	-0.458079	0.582535	-0.584984	Putative disease resistance RPPI3-like protein 1
AH12G02120	Chr12	2641657	G	C	S	13	12.1	0.92	S	19	31.6	0.16	0.77	NON_SYNONYMOUS_CODING	gCc/gGc	0.456806	-0.458062	0.582519	-0.584957	Putative disease resistance RPPI3-like protein 1
AH12G02130	Chr12	2651310	C	T	Y	13	6.7	0.54	Y	11	8.3	0.27	0.27	NON_SYNONYMOUS_CODING	gGu/gAt	0.456798	-0.458054	0.582513	-0.584944	Putative disease resistance RPPI3-like protein 1
AH12G02130	Chr12	2651350	A	C	M	7	4.3	0.43	M	9	7.2	0.22	0.21	NON_SYNONYMOUS_CODING	Ttg/Gtg	0.456798	-0.458054	0.582513	-0.584944	Putative disease resistance RPPI3-like protein 1
AH12G02130	Chr12	2651430	T	A	W	9	4.5	0.56	T	10	9.1	0.10	0.46	NON_SYNONYMOUS_CODING	aAt/aTt	0.456798	-0.458054	0.582513	-0.584944	Putative disease resistance RPPI3-like protein 1
AH12G02310	Chr12	2777261	G	T	K	11	9.2	0.82	T	14	0.14	0.00	0.82	NON_SYNONYMOUS_CODING	agC/agT	0.456652	-0.457914	0.582398	-0.584766	Putative disease resistance RPPI3-like protein 1
AH12G02310	Chr12	2816492	T	G	K	17	13.4	0.76	K	15	2.13	0.13	0.63	STOP_LOST	tAa/tCa	0.456653	-0.457913	0.582412	-0.582412	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820701	G	A	G	23	23.0	1.00	R	12	11.1	0.92	0.08	NON_SYNONYMOUS_CODING	aGg/aAg	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820718	A	C	A	24	24.0	1.00	M	13	11.2	0.85	0.15	NON_SYNONYMOUS_CODING	Atc/Ctc	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820776	A	G	A	28	28.0	1.00	R	15	12.3	0.80	0.20	NON_SYNONYMOUS_CODING	gAa/gGa	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820802	A	C	A	27	27.0	1.00	M	16	12.4	0.75	0.25	NON_SYNONYMOUS_CODING	Att/Ctt	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820833	G	C	G	28	28.0	1.00	S	15	11.4	0.73	0.27	NON_SYNONYMOUS_CODING	gGu/gCt	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820846	T	A	T	26	26.0	1.00	W	14	11.3	0.79	0.21	NON_SYNONYMOUS_CODING	agT/agA	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2820944	G	A	G	22	22.0	1.00	R	11	5.6	0.45	0.55	NON_SYNONYMOUS_CODING	cGg/cAg	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2822450	A	C	M	23	20.3	0.87	M	23	4.19	0.17	0.70	NON_SYNONYMOUS_CODING	gAu/gCt	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02330	Chr12	2822465	A	G	R	21	18.3	0.86	R	22	5.17	0.23	0.63	NON_SYNONYMOUS_CODING	cAu/cCt	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB-SNP-index-SB-SNP-index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G02370	Chr12	2823195	T	C	T	19	19.0	1.00	Y	6	2.4	0.33	0.67	NON_SYNONYMOUS_CODING	tAu/Gt	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02370	Chr12	2823222	G	A	G	20	20.0	1.00	R	7	2.5	0.29	0.71	NON_SYNONYMOUS_CODING	Gca/Cca	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02370	Chr12	2823262	G	C	G	26	26.0	1.00	S	5	1.4	0.20	0.80	NON_SYNONYMOUS_CODING	Gaa/Caa	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02370	Chr12	2823268	C	A	C	24	24.0	1.00	M	6	1.5	0.17	0.83	NON_SYNONYMOUS_CODING	Ctg/Atg	0.456665	-0.457925	0.582427	-0.584779	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2845356	C	A	M	19	14.5	0.74	M	19	4.15	0.21	0.53	NON_SYNONYMOUS_CODING	Ctt/Att	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2845498	G	A	R	27	25.2	0.93	G	25	23.2	0.92	0.01	NON_SYNONYMOUS_CODING	cGc/cAc	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2845513	A	G	R	26	24.2	0.92	A	25	24.1	0.96	-0.04	NON_SYNONYMOUS_CODING	aAg/aGg	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2845530	T	C	Y	27	25.2	0.93	T	25	24.1	0.96	-0.03	NON_SYNONYMOUS_CODING	Tat/Cat	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2846684	G	T	G	22	22.0	1.00	K	18	10.8	0.56	0.44	NON_SYNONYMOUS_CODING	agC/agT	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2846808	A	G	A	17	17.0	1.00	R	25	13.12	0.52	0.48	NON_SYNONYMOUS_CODING	Agg/Ggg	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2846850	T	A	T	15	15.0	1.00	W	26	15.11	0.58	0.42	NON_SYNONYMOUS_CODING	Tcc/ Acc	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2846856	C	G	C	16	16.0	1.00	S	25	14.11	0.56	0.44	NON_SYNONYMOUS_CODING	Caa/ Gaa	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2846949	A	G	A	11	11.0	1.00	R	15	12.3	0.80	0.20	NON_SYNONYMOUS_CODING	Aaa/ Gaa	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02390	Chr12	2847652	G	T	G	15	14.1	0.93	K	6	2.4	0.33	0.60	NON_SYNONYMOUS_CODING	cGc/cTc	0.456684	-0.457942	0.582449	-0.5848	Putative disease resistance RPPI3-like protein 1
AH12G02410	Chr12	2866827	A	G	A	21	21.0	1.00	R	4	1.3	0.25	0.75	NON_SYNONYMOUS_CODING	Aag/ Gag	0.456704	-0.457961	0.582471	-0.584825	Putative disease resistance RPPI3-like protein 1
AH12G02410	Chr12	2867664	G	C	G	18	17.1	0.94	C	19	1.18	0.05	0.89	NON_SYNONYMOUS_CODING	Gac/ Cac	0.456704	-0.457961	0.582471	-0.584825	Putative disease resistance RPPI3-like protein 1
AH12G02410	Chr12	2868121	C	G	S	25	21.4	0.84	S	17	2.15	0.12	0.72	NON_SYNONYMOUS_CODING	cGc/ cGc	0.456704	-0.457961	0.582471	-0.584825	Putative disease resistance RPPI3-like protein 1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB-SNP-index-SB-SNP-index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G02410	Chr12	2868129	C	G	S	27	22.5	0.81	S	17	3.14	0.18	0.64	NON_SYNONYMOUS_CODING	Caa/Gaa	0.456704	-0.457961	0.582471	-0.584825	Putative disease resistance RPP13-like protein 1
AH12G02520	Chr12	2952570	C	T	Y	24	21.3	0.88	Y	18	3.15	0.17	0.71	NON_SYNONYMOUS_CODING	Gga/Agg	0.456781	-0.458031	0.582557	-0.584916	1-aminocyclopropane-1-carboxylate synthase-like protein 1
AH12G02650	Chr12	3067519	T	C	Y	26	21.5	0.81	Y	12	2.10	0.17	0.64	NON_SYNONYMOUS_CODING	rTg/rCg	0.456889	-0.458138	0.582693	-0.585057	Reticon-like protein B1
AH12G02880	Chr12	3276550	G	C	G	16	15.1	0.94	S	21	3.18	0.14	0.79	NON_SYNONYMOUS_CODING	tGg/tGc	0.457055	-0.458315	0.582914	-0.58527	Putative disease resistance RPP13-like protein 1
AH12G02960	Chr12	3347238	C	A	C	17	17.0	1.00	M	7	4.3	0.57	0.43	NON_SYNONYMOUS_CODING	Cgt/Agg	0.457096	-0.458361	0.582977	-0.585315	Transmembrane receptors 3B1P binding
AH12G02960	Chr12	3347384	T	A	T	27	25.2	0.93	W	8	4.4	0.50	0.43	NON_SYNONYMOUS_CODING	ttT/tTA	0.457096	-0.458361	0.582977	-0.585315	Transmembrane receptors 3B1P binding
AH12G02960	Chr12	3348195	C	G	C	5	5.0	1.00	S	10	1.9	0.10	0.90	NON_SYNONYMOUS_CODING	Cat/Cat	0.457096	-0.458361	0.582977	-0.585315	Transmembrane receptors 3B1P binding
AH12G03000	Chr12	3392506	C	G	C	24	24.0	1.00	S	9	2.7	0.22	0.78	NON_SYNONYMOUS_CODING	Ctt/Gtt	0.457119	-0.458386	0.583012	-0.585331	Protein of unknown function (DUF594)
AH12G03000	Chr12	3392515	C	A	C	22	21.1	0.95	M	9	2.7	0.22	0.73	NON_SYNONYMOUS_CODING	Ctg/Atg	0.457119	-0.458386	0.583012	-0.585331	Protein of unknown function (DUF595)
AH12G03240	Chr12	3693638	C	G	C	13	13.0	1	S	11	1.10	0.09	0.91	NON_SYNONYMOUS_CODING	aCt/aGt	0.457106	-0.4584	0.583117	-0.585359	putative disease resistance RPP13-like protein 1
AH12G03240	Chr12	3715337	G	C	G	13	13.0	1.00	S	22	4.18	0.18	0.82	NON_SYNONYMOUS_CODING	Gat/Cat	0.457108	-0.458403	0.583131	-0.585372	BAG family molecular chaperone regulator 6
AH12G03240	Chr12	3715406	T	C	T	17	17.0	1.00	Y	15	4.11	0.27	0.73	NON_SYNONYMOUS_CODING	Tgt/Cgt	0.457108	-0.458403	0.583131	-0.585372	BAG family molecular chaperone regulator 6
AH12G03240	Chr12	3715462	T	G	T	18	18.0	1.00	K	5	1.4	0.20	0.80	NON_SYNONYMOUS_CODING	atT/atG	0.457108	-0.458403	0.583131	-0.585372	BAG family molecular chaperone regulator 6
AH12G03290	Chr12	3745691	G	A	R	21	18.3	0.86	R	13	2.11	0.15	0.70	NON_SYNONYMOUS_CODING	cGt/cAt	0.457108	-0.458406	0.583152	-0.585391	Serine/threonine-protein phosphatase 7
AH12G03300	Chr12	3748516	C	T	Y	18	14.4	0.78	Y	20	4.16	0.20	0.58	NON_SYNONYMOUS_CODING	cGat/cAa	0.457108	-0.458406	0.583152	-0.585391	Aminotransferase-like 2C plant mobile domain family protein
AH12G03300	Chr12	3998818	G	A	G	20	20.0	1.00	R	14	5.9	0.36	0.64	NON_SYNONYMOUS_CODING	Ggt/Agg	0.456817	-0.45814	0.582958	-0.585158	MuDR family transposase

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB index-SB index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G03500	Chr12	3999460	G	T	K	25	22.3	0.88	K	17	4.13	0.24	0.64	NON_SYNONYMOUS_CODING	aGg/aTg	0.456817	-0.45814	0.582958	-0.585158	MaDR family transposase
AH12G03500	Chr12	3999687	A	G	R	22	20.2	0.91	G	17	0.17	0.00	0.91	NON_SYNONYMOUS_CODING	Aac/Gac	0.46	-0.46	0.58	-0.59	MaDR family transposase
AH12G03560	Chr12	4096307	A	T	A	15	15.0	1.00	T	5	0.5	0.00	1.00	NON_SYNONYMOUS_CODING	gTg/gAg	0.46	-0.46	0.58	-0.59	UDP-glycosyltransferase 72B1
AH12G03600	Chr12	4142557	T	G	K	17	15.2	0.88	K	7	1.6	0.14	0.74	NON_SYNONYMOUS_CODING	TcT/Gat	0.46	-0.46	0.58	-0.59	Disease resistance protein TAO1
AH12G03600	Chr12	4142566	G	A	R	17	15.2	0.88	A	5	0.5	0.00	0.88	NON_SYNONYMOUS_CODING	Gtt/Att	0.46	-0.46	0.58	-0.59	Disease resistance protein TAO1
AH12G03600	Chr12	4142572	G	A	R	17	15.2	0.88	A	5	0.5	0.00	0.88	NON_SYNONYMOUS_CODING	Ggt/AgT	0.46	-0.46	0.58	-0.59	Disease resistance protein TAO1
AH12G03900	Chr12	4633134	C	T	Y	19	17.2	0.89	Y	15	2.13	0.13	0.76	NON_SYNONYMOUS_CODING	Ggt/Gcg/ Acg	0.46	-0.46	0.58	-0.58	Cytochrome b561 and DOMON domain-containing protein
AH12G03980	Chr12	4691788	C	T	Y	20	16.4	0.8	Y	17	4.13	0.24	0.56	STOP_GAINED	Caa/ Taa	0.456446	-0.457759	0.582353	-0.585012	homolog of histone chaperone HIRA
AH12G03980	Chr12	4696605	G	A	G	25	25.0	1.00	R	12	2.10	0.17	0.83	NON_SYNONYMOUS_CODING	Gtt/Att	0.456446	-0.457759	0.582353	-0.585012	homolog of histone chaperone HIRA
AH12G04010	Chr12	4718691	G	T	K	21	16.5	0.76	T	16	1.15	0.06	0.70	NON_SYNONYMOUS_CODING	Ctg/Atg	0.456475	-0.457785	0.582373	-0.585056	Eukaryotic aspartyl protease family protein
AH12G04170	Chr12	5007281	T	A	T	23	23.0	1.00	W	22	19.3	0.86	0.14	NON_SYNONYMOUS_CODING	cAc/ cTc	0.457503	-0.45874	0.58321	-0.586473	Endosomal targeting BRO1-like domain-containing protein
AH12G04170	Chr12	5007411	A	C	A	26	26.0	1.00	M	22	17.5	0.77	0.23	NON_SYNONYMOUS_CODING	Tca/ Gca	0.457503	-0.45874	0.58321	-0.586473	Endosomal targeting BRO1-like domain-containing protein
AH12G04170	Chr12	5007427	G	C	G	27	27.0	1.00	S	22	17.5	0.77	0.23	NON_SYNONYMOUS_CODING	ttC/ttG	0.457503	-0.45874	0.58321	-0.586473	Endosomal targeting BRO1-like domain-containing protein
AH12G04170	Chr12	5007448	C	G	C	24	24.0	1.00	S	18	15.3	0.83	0.17	NON_SYNONYMOUS_CODING	tgG/ tgC	0.457503	-0.45874	0.58321	-0.586473	Endosomal targeting BRO1-like domain-containing protein
AH12G04170	Chr12	5007454	C	G	C	27	27.0	1.00	S	17	14.3	0.82	0.18	NON_SYNONYMOUS_CODING	ttG/ttC	0.457503	-0.45874	0.58321	-0.586473	Endosomal targeting BRO1-like domain-containing protein
AH12G04330	Chr12	5187080	C	T	C	20	18.2	0.90	Y	10	3.7	0.30	0.60	NON_SYNONYMOUS_CODING	Gag/ Aag	0.458456	-0.459622	0.584098	-0.587787	P-loop containing nucleoside triphosphate hydrolases
AH12G04490	Chr12	5526242	C	T	Y	25	19.6	0.76	Y	14	4.10	0.29	0.47	NON_SYNONYMOUS_CODING	gGat/ gAa	0.460031	-0.461049	0.585491	-0.5901	superfamily protein Uncharacterized protein

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB index-SB index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G04580	Chr12	5765500	G	T	G	18	17.1	0.94	K	6	1.5	0.17	0.78	NON_SYNONYMOUS_CODING	aaC/aaA	0.460793	-0.461707	0.58614	-0.591388	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AH12G04670	Chr12	5864398	G	T	K	21	17.4	0.81	K	14	3.11	0.21	0.60	NON_SYNONYMOUS_CODING	GgT/Tgt	0.46102	-0.461893	0.586301	-0.591789	Uncharacterized protein
AH12G04700	Chr12	5914144	G	A	G	18	17.1	0.94	R	13	1.12	0.08	0.87	NON_SYNONYMOUS_CODING	Gtg/Atg	0.461224	-0.462076	0.586503	-0.592101	Uncharacterized protein
AH12G04740	Chr12	6018424	T	C	Y	24	17.7	0.71	Y	16	5.11	0.31	0.40	NON_SYNONYMOUS_CODING	Tcc/Ccc	0.461787	-0.462593	0.587129	-0.59291	3-ketacyl-CoA synthase 1
AH12G04980	Chr12	6321518	C	T	Y	22	17.5	0.77	Y	14	3.11	0.21	0.56	NON_SYNONYMOUS_CODING	cCc/cTc	0.462862	-0.46355	0.588336	-0.594617	Uncharacterized protein
AH12G05040	Chr12	6391830	A	G	R	19	15.4	0.79	R	18	3.15	0.17	0.62	NON_SYNONYMOUS_CODING	cAg/cGg	0.463056	-0.463716	0.588597	-0.594957	Pyruvate dehydrogenase E1 component subunit alpha-3/2C chloroplastic
AH12G05080	Chr12	6517918	T	C	T	12	12.0	1.00	Y	9	8.1	0.89	0.11	NON_SYNONYMOUS_CODING	Ttc/Ctc	0.463369	-0.463988	0.589041	-0.595523	Probably inactive leucine-rich repeat receptor-like protein kinase
AH12G05080	Chr12	6517937	A	T	A	14	14.0	1.00	W	12	10.2	0.83	0.17	NON_SYNONYMOUS_CODING	gAc/gTc	0.463369	-0.463988	0.589041	-0.595523	Probably inactive leucine-rich repeat receptor-like protein kinase
AH12G05080	Chr12	6517943	C	A	C	15	15.0	1.00	M	12	10.2	0.83	0.17	NON_SYNONYMOUS_CODING	tC/tAt	0.463369	-0.463988	0.589041	-0.595523	Probably inactive leucine-rich repeat receptor-like protein kinase
AH12G05080	Chr12	6518088	T	A	T	18	18.0	1.00	W	7	5.2	0.71	0.29	NON_SYNONYMOUS_CODING	ttT/tTA	0.463369	-0.463988	0.589041	-0.595523	Probably inactive leucine-rich repeat receptor-like protein kinase
AH12G05100	Chr12	6594841	G	A	R	17	13.4	0.76	R	21	3.18	0.14	0.62	NON_SYNONYMOUS_CODING	gCc/gTg	0.463457	-0.464051	0.589192	-0.595747	Probably inactive leucine-rich repeat receptor-like protein kinase
AH12G05120	Chr12	6605230	G	A	R	22	17.5	0.77	A	13	0.13	0.00	0.77	NON_SYNONYMOUS_CODING	aCa/aTa	0.463458	-0.46405	0.589199	-0.595763	Flavone 3'-O-methyltransferase 1
AH12G05120	Chr12	6606476	T	A	T	20	20.0	1.00	W	13	10.3	0.77	0.23	NON_SYNONYMOUS_CODING	Atc/Ttc	0.463458	-0.46405	0.589199	-0.595763	Flavone 3'-O-methyltransferase 1
AH12G05120	Chr12	6607843	G	C	G	25	23.2	0.92	S	14	10.4	0.71	0.21	NON_SYNONYMOUS_CODING	aCr/aGt	0.463458	-0.46405	0.589199	-0.595763	Flavone 3'-O-methyltransferase 1
AH12G05140	Chr12	6655924	G	A	R	15	13.2	0.87	R	10	3.7	0.30	0.57	NON_SYNONYMOUS_CODING	gCt/gTt	0.463453	-0.464034	0.58922	-0.595823	Flavone 3'-O-methyltransferase 1
AH12G05250	Chr12	6843032	T	G	K	14	13.1	0.93	G	9	0.9	0.00	0.93	NON_SYNONYMOUS_CODING	Tcc/Gcc	0.463323	-0.463869	0.589187	-0.59586	B3 domain-containing mRNA1on factor NGA1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index (RB index-SB index)	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12G05320	Chr12	6989102	A	G	R	10	9.1	0.90	R	12	2.10	0.17	0.73	NON_SYNONYMOUS_CODING	Tcc/Ccc	0.462905	-0.463448	0.588882	-0.59155	Serine/threonine-protein phosphatase 7
AH12G05420	Chr12	7082034	T	C	T	16	16.0	1.00	Y	15	1.14	0.07	0.93	NON_SYNONYMOUS_CODING	Aca/Gca	0.462547	-0.463075	0.588567	-0.595158	Probable inactive purple acid phosphatase 29
AH12G05820	Chr12	7671331	C	G	C	23	22.1	0.96	S	7	1.6	0.14	0.81	NON_SYNONYMOUS_CODING	caC/caG	0.460651	-0.461098	0.586887	-0.593277	FRGIDA-like protein 3
AH12G05970	Chr12	7992939	G	A	R	12	11.1	0.92	R	15	2.13	0.13	0.78	NON_SYNONYMOUS_CODING	Cgc/Tgc	0.459751	-0.460171	0.585875	-0.592182	Aminotransferase-like 2C plant mobile domain family protein
AH12G06150	Chr12	8202579	C	T	T	21	0.21	1.00	Y	17	13.4	0.24	0.76	NON_SYNONYMOUS_CODING	Gaa/Aaa	0.459288	-0.459699	0.585408	-0.591564	AT-rich interactive domain-containing protein 4
AH12G06180	Chr12	8256014	C	A	C	21	20.1	0.95	M	21	3.18	0.14	0.81	NON_SYNONYMOUS_CODING	CaT/Aat	0.45923	-0.45964	0.585372	-0.591479	RING/U-box superfamily protein
AH12G06180	Chr12	8260084	G	A	G	22	22.0	1.00	R	9	6.3	0.67	0.33	NON_SYNONYMOUS_CODING	aGg/aAg	0.459218	-0.459627	0.585364	-0.591462	RING/U-box superfamily protein
AH12G06180	Chr12	8260159	T	C	T	16	16.0	1.00	Y	14	4.10	0.29	0.71	NON_SYNONYMOUS_CODING	tTt/tCt	0.459218	-0.459627	0.585364	-0.591462	RING/U-box superfamily protein
AH12G06180	Chr12	8260203	G	A	G	18	18.0	1.00	R	24	15.9	0.63	0.38	NON_SYNONYMOUS_CODING	Gtt/Att	0.459218	-0.459627	0.585364	-0.591462	RING/U-box superfamily protein
AH12G06180	Chr12	8271833	A	G	A	20	19.1	0.95	R	18	6.12	0.33	0.62	NON_SYNONYMOUS_CODING	tAt/tGt	0.459205	-0.459615	0.585355	-0.591445	RING/U-box superfamily protein
AH12G06180	Chr12	8271850	G	C	G	22	21.1	0.95	S	18	6.12	0.33	0.62	NON_SYNONYMOUS_CODING	Gaa/Caa	0.459205	-0.459615	0.585355	-0.591445	RING/U-box superfamily protein
AH12G06210	Chr12	8303749	A	G	A	26	26.0	1.00	R	6	5.1	0.83	0.17	NON_SYNONYMOUS_CODING	Aca/Gca	0.459187	-0.459596	0.585344	-0.591416	RING/U-box superfamily protein
AH12G06210	Chr12	8303776	A	G	A	23	23.0	1.00	R	6	3.3	0.50	0.50	NON_SYNONYMOUS_CODING	Aaa/Gaa	0.459187	-0.459596	0.585344	-0.591416	RING/U-box superfamily protein
AH12G06250	Chr12	8416831	T	G	T	16	16.0	1.00	K	6	1.5	0.17	0.83	NON_SYNONYMOUS_CODING	Tgt/Ggt	0.459105	-0.459511	0.585247	-0.591281	RING/U-box superfamily protein
AH12G06300	Chr12	8578649	C	T	Y	16	13.3	0.81	Y	16	3.13	0.19	0.63	NON_SYNONYMOUS_CODING	gCt/gTt	0.458995	-0.459395	0.585101	-0.591125	Exocyst subunit exo70 family protein B2
AH12G06300	Chr12	8579347	C	T	C	12	12.0	1.00	Y	18	3.15	0.17	0.83	NON_SYNONYMOUS_CODING	Cgg/Tgg	0.458995	-0.459395	0.585101	-0.591125	Exocyst subunit exo70 family protein B2
AH12G06320	Chr12	8636540	G	T	G	18	18.0	1.00	T	9	0.9	0.00	1.00	NON_SYNONYMOUS_CODING	aaC/aaa	0.458954	-0.459351	0.585052	-0.591055	Putative disease resistance RPP13-like protein 1
AH12G06320	Chr12	8636563	C	A	C	18	18.0	1.00	A	8	0.8	0.00	1.00	NON_SYNONYMOUS_CODING	Gat/Tat	0.458954	-0.459351	0.585052	-0.591055	Putative disease resistance RPP13-like protein 1
AH12G06320	Chr12	8636575	G	A	G	13	13.0	1.00	A	7	0.7	0.00	1.00	NON_SYNONYMOUS_CODING	Cac/Tac	0.458954	-0.459351	0.585052	-0.591055	Putative disease resistance RPP13-like protein 1

(Continued)

TABLE 2 Continued

Gene ID	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref. Alt	SNP-index of RB	Susceptible bulk (SB) base	Number of reads covering the site (X coverage) in susceptible bulk (SB)	SB Depths of Ref. Alt	SNP-index of SB	Delta SNP-index	SNP substitution effect	Amino acid change	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene function
AH12C06430	Chr12	8714663	C	A	M	12	11.1	0.92	M	19	3.16	0.16	0.76	NON_SYNONYMOUS_CODING	Cta/Ata	0.45886	-0.459253	0.584957	-0.590896	Duplicated homeodomain-like superfamily protein
AH12C06450	Chr12	8735310	C	T	C	14	14.0	1.00	Y	11	2.9	0.18	0.82	NON_SYNONYMOUS_CODING	aGaa/aAa	0.458829	-0.459221	0.584923	-0.590847	Endosomal targeting BRO1-like domain-containing protein
AH12C06480	Chr12	8738727	G	T	G	16	15.1	0.94	K	9	4.5	0.44	0.49	NON_SYNONYMOUS_CODING	aCaa/aAa	0.458799	-0.45919	0.58489	-0.5908	GDSL esterase/lipase 5
AH12C06480	Chr12	8767570	C	G	C	18	18.0	1.00	S	9	2.7	0.22	0.78	NON_SYNONYMOUS_CODING	caG/caC	0.458783	-0.459174	0.584873	-0.590776	GDSL esterase/lipase 5
AH12C06900	Chr12	8792582	T	G	T	21	21.0	1.00	K	13	7.6	0.54	0.46	NON_SYNONYMOUS_CODING	Tta/Gta	0.458733	-0.459123	0.584819	-0.590701	GDSL esterase/lipase 3
AH12C06560	Chr12	8898415	T	G	G	20	21.8	0.90	K	26	23.3	0.12	0.78	NON_SYNONYMOUS_CODING	gAa/gCt	0.458603	-0.458987	0.584691	-0.590474	UDP-glycosyltransferase 72B3
AH12C06950	Chr12	9324156	C	G	C	20	19.1	0.95	S	12	8.4	0.67	0.28	NON_SYNONYMOUS_CODING	tGg/tCg	0.458109	-0.458489	0.584069	-0.590328	Pentatricopeptide repeat-containing protein
AH12C06950	Chr12	9324470	T	C	T	16	15.1	0.94	Y	18	11.7	0.61	0.33	NON_SYNONYMOUS_CODING	aAa/aGt	0.458109	-0.458489	0.584069	-0.590328	Pentatricopeptide repeat-containing protein
AH12C07020	Chr12	9478950	G	A	R	8	1.7	0.88	R	6	5.1	0.17	0.71	NON_SYNONYMOUS_CODING	tGt/aAc	0.458059	-0.458444	0.583973	-0.590016	Squamosa promoter-binding-like protein 8
AH12C07090	Chr12	9663318	G	T	K	24	21.3	0.88	K	29	8.21	0.28	0.60	NON_SYNONYMOUS_CODING	Gaa/Taa	0.457966	-0.458357	0.58382	-0.588646	mRNAon regulators

TABLE 3 Identification of InDels in putative candidate genes in the genomic region for resistance to on chromosome 12.

Gene	Chromosome	Physical position (bp)	Reference Genome	Alternative site	Resistant bulk (RB) base	Number of reads covering the site (X coverage) in resistant bulk (RB)	RB Depths of Ref, Alt	SNP-index of RB	Susceptible bulk(SB)base	Number of reads covering the site (X coverage) in susceptible bulk(SB)	SB Depths of Ref, Alt	SNP-index of SB	delta SNP-index (RB SNP-index-SB SNP-index)	SNP substitution effect	U95 (95% confidence interval upper side)	L95 (95% confidence interval lower side)	U99 (99% confidence interval upper side)	L99 (99% confidence interval lower side)	Gene Function
AH12G01550	Chr12	1888286	TCC	T	TCC	23	22,1	0.96	TCC,T	9	7,2	0.78	0.18	FRAME_SHIFT	0.499429	-0.498581	0.630045	-0.613463	putative disease resistance protein At3g14460 isoform X2 [Arachis ipaensis]
AH12G01670	Chr12	2019127	CCTATTCTACG CCTCACAAAATCT	C	CCTATTCTACG CCTCACAAAATCT	28	26,2	0.93	CCTATTCTACGCTCACAAAATCT, C	22	17,5	0.77	0.16	FRAME_SHIFT	0.499034	-0.498174	0.629563	-0.613179	HXXXD-type acyl-transferase family protein
AH12G01670	Chr12	2020084	T	TGGTGAAACA	T	21	21,0	1.00	T,TGGTGAAACA	15	10,5	0.67	0.33	CODON_INSERTION	0.498917	-0.498054	0.629416	-0.613083	HXXXD-type acyl-transferase family protein
AH12G01920	Chr12	2473096	A	ACGTTT	A,ACGTTT	22	19,3	0.86	A,ACGTTT	16	2,14	0.13	0.74	FRAME_SHIFT	0.496175	-0.495264	0.626054	-0.611224	Putative disease resistance protein At3g14460
AH12G01980	Chr12	2534345	CTCCGTACCAAG	C	C,CTCCGTACCAAG	24	22,2	0.92	C,CTCCGTACCAAG	20	18,2	0.90	0.02	FRAME_SHIFT	0.495902	-0.494987	0.625693	-0.611031	Putative disease resistance RPP13-like protein 1
AH12G01980	Chr12	2539604	TA	T	TA	24	24,0	1.00	TA,T	20	17,3	0.85	0.15	FRAME_SHIFT	0.495902	-0.494987	0.625693	-0.611031	Putative disease resistance RPP13-like protein 1
AH12G01980	Chr12	2539716	C	CCA	C	25	25,0	1.00	CCA,C	25	15,10	0.60	0.40	FRAME_SHIFT	0.495902	-0.494987	0.625693	-0.611031	Putative disease resistance RPP13-like protein 1
AH12G02090	Chr12	2620516	TGG	T	TGG	24	24,0	1.00	T,TGG	9	8,1	0.89	0.11	FRAME_SHIFT	0.495698	-0.494779	0.625463	-0.610928	Putative disease resistance RPP13-like protein 1
AH12G02390	Chr12	2847658	T	TCAAA	T	16	15,1	0.94	T,TCAAA	6	2,4	0.33	0.60	FRAME_SHIFT	0.495495	-0.494571	0.625259	-0.610874	Putative disease resistance RPP13-like protein 1
AH12G02440	Chr12	2879448	T	TGAG	T	11	11,0	1.00	T,TGAG	6	5,1	0.83	0.17	CODON_CHANGE_PLUS_CODON_INSERTION	0.495507	-0.494586	0.625287	-0.610914	Putative disease resistance RPP13-like protein 1
AH12G02600	Chr12	3030085	T	TGATGGTGAGA CTTTGAAGAGCA ACCAAGTCCTTC TCTTGAAATGAC ACAAATGAGTTGCAGTGATG	TGATGGTGAGACTTT GAAGAGCAACCAAGTC CTTCTCTTGGAATGA CACAAATGAGTTGCAGTGATG, T	8	7,1	0.88	T,TGATGGTGAG ACITTTGAAGAGCA ACCAAGTCCTTCTC TTGGAAATGACAC AAATGAGTTGCAGTGATG	13	2,11	0.15	0.72	CODON_INSERTION	0.495575	-0.494661	0.625474	-0.61112	Putative disease resistance RPP13-like protein 1
AH12G03240	Chr12	3715392	ATCG	A	ATCG	17	17,0	1.00	A,ATCG	17	4,13	0.24	0.76	CODON_DELETION	0.496363	-0.494543	0.627006	-0.612884	BAG family molecular chaperone regulator 6
AH12G06110	Chr12	8165877	G	GGATTAGTGGTGCAGCAGCTTGT	G	20	20,0	1.00	G, GGATTAGTGGTGCAGCAGCTTGT	10	3,7	0.30	0.70	FRAME_SHIFT	0.502752	-0.502787	0.64146	-0.630466	hypothetical protein MTR_5g086360
AH12G06180	Chr12	8289562	CTAA	C	CTAA	10	10,0	1.00	CTAA,C	21	19,2	0.90	0.10	CODON_DELETION	0.502481	-0.502506	0.641028	-0.630118	E3 ubiquitin-protein ligase RNF144A-like

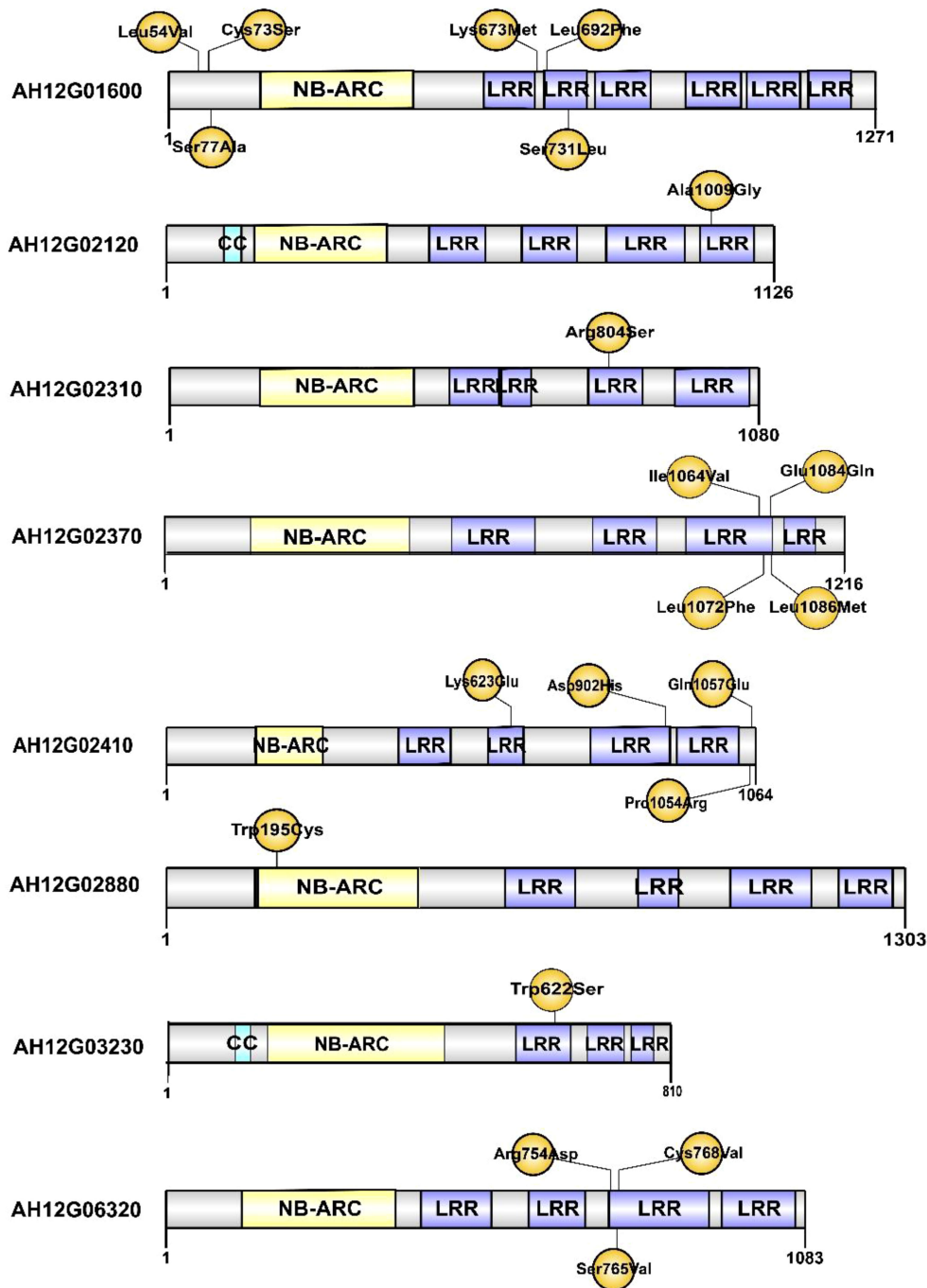


FIGURE 4
The putative resistance-related proteins associated with resistance to *Ralstonia solanacearum* infections. CC: Coil coiled. NB-ARC: Nucleotide-binding adapter shared by APAF-1 R proteins and CED-4. LRR: Leucine-rich repeat domain. The positions of amino acid changes caused by nonsynonymous SNPs are shown in yellow.

seq approach is an increasingly popular sequencing-based method for the identification of candidate genomic regions associated with target traits in peanut (Varshney et al., 2019; Pandey et al., 2020). As it only requires whole genome sequences

of parents and extreme trait bulks from the mapping population, it is economical, efficient, fast, and cost-effective (Takagi et al., 2013). Traditional QTL mapping methods are limited in the fine mapping of target genes and QTLs because they lack both high-

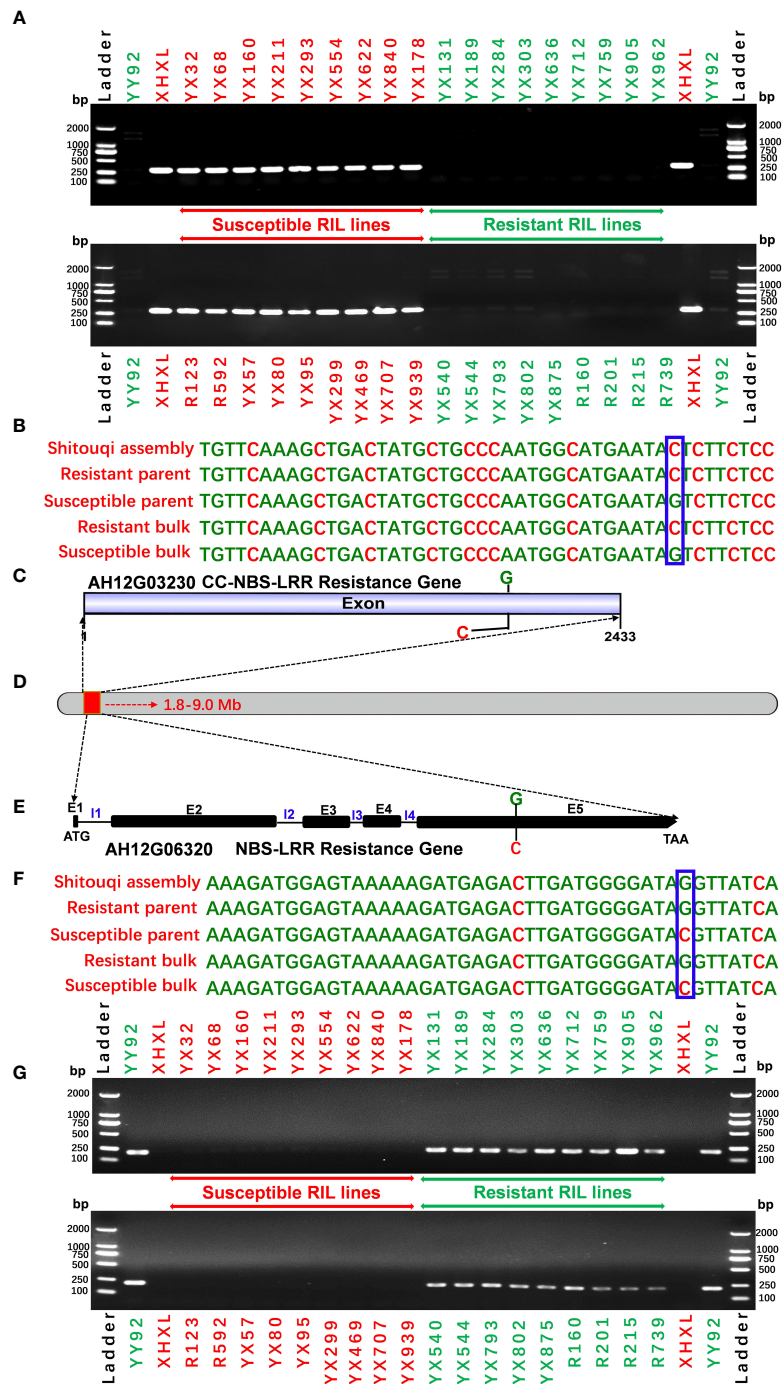


FIGURE 5

Validation of putative candidate gene-based markers of bacterial wilt resistance. **(A)** The SNP marker validation of candidate gene *AH12G03230* using a validation set comprising the resistant parent YY92, susceptible parent XHXL, and susceptible and resistant RIL lines. **(B)** SNP variation in the *AH12G03230* gene. **(C)** The *AH12G03230* gene is predicted to encode the CC-NBS-LRR resistance protein. **(D)** Putative genomic region on Chromosome 12 of *Arachis hypogaea* that encodes resistance to *Ralstonia solanacearum* infections. **(E)** The *AH12G06320* gene is predicted to encode the NBS-LRR resistant protein (E1 to E5 refer to exon numbers while I1 to I4 refer to intron numbers). **(F)** SNP variation in the *AH12G06000* gene and **(G)** marker validation on a validation set comprising resistant parent YY92, susceptible parent XHXL, susceptible RIL lines (YX32, YX68, YX160, YX211, YX293, YX554, YX622, YX840 and YX178, R123, R592, YX57, YX80, YX95, YX299, YX469, YX707 and YX939), and resistant RIL lines (YX131, YX189, YX284, YX303, YX636, YX712, YX759, YX905, YX962, YX540, YX544, YX793, YX802, YX875, R160, R201, R215 and R739).

density genetic maps and a series of near-isogenic lines (Pandey et al., 2017). Despite not having a large segregation population as a prerequisite, the QTL-seq approach was successful in identifying candidate genes for many crop traits (Das et al., 2014; Lu et al., 2014; Illa et al., 2015; Singh et al., 2016; Wang et al., 2016; Pandey et al., 2017; Srivastava et al., 2017; Chen et al., 2018; Clevenger et al., 2018; Luo et al., 2019; Bommisetty et al., 2020; Kumar et al., 2020; Lei et al., 2020; Tudor et al., 2020; Zhao et al., 2020; Cao et al., 2021; Dong et al., 2021; Topcu et al., 2021; Wang et al., 2021; Yang et al., 2021; Zhang et al., 2021). In the present study, a QTL-seq approach was successfully applied to identify genomic regions and candidate genes for RRS using resequencing data of both parental genotypes and pooled samples of the RIL population (Yueyou 92×Xinhuixiaoli) (Supplementary Figure 1), which corroborated our previously reported QTL mapping findings (Zhao et al., 2016).

The use of a common reference genome that is associated with deep sequencing and large bulks should result in highly accurate maps. As per the original QTL-seq approach (Takagi et al., 2013), the genome assemblies of either one or both parents were used as reference to analyze the SNP variants in the two extreme bulks based on diploid reference genomes due to the unavailability of the cultivated peanut genome (Pandey et al., 2017; Luo et al., 2019; Kumar et al., 2020; Chen et al., 2021). Candidate genomic regions were then associated with target traits by the Δ SNP index method (Takagi et al., 2013). The choice of the parental reference genome possibly affects this association of candidate genomic regions due to differing levels of alignment errors (Luo et al., 2019). Moreover, the algorithm uses the reference genomes to replace the parental genomes in bigger diversity areas, which may cause erroneous assemblies of the parent genomes, especially for wild diploid ancestors (unpublished data). To increase the reliability of identified genomic regions and candidate genes, we used the *Arachis hypogaea* Shitouqi genome as a reference. Shitouqi (*A. h. fastigiata* var. *vulgaris*), belongs to the subsp. *fastigiata*, as does the parents of the population, and its high-quality genome sequence was recently reported (Zhuang et al., 2019). Based on a large RIL population of 581 individual lines from the cross of resistant YY92 and susceptible Xinhuixiaoli, 30 resistant and 30 susceptible lines were selected to respectively construct the extreme R and S-bulks (Figure 1 and Supplementary Table 1). We generated 108.38 and 96.92 Gb of sequence reads at a sequencing depth of 38× and 34× for the R- and S-bulks, respectively (Table 1). Nearly 98% of these reads were correctly mapped onto the reference genome. High densities of homozygous SNPs (381,642) and InDels (98,918) between parents were identified by resequencing for RRS mapping (Supplementary Tables 2, 3). The candidate region of 5.73 Mb on Chr12 was identified by combining the ED and Δ SNP/InDel index algorithms for both SNPs and InDels (Figures 2, 5; Table 2; Supplementary Tables 2, 3) at a $P < 0.01$ confidence level. These aided the precise and accurate discovery of candidate genomic regions, genes, and SNPs/Indels markers associated with RRS.

The clear extreme phenotypic differences between parents as well as those of the pooled population were the crucial prerequisite for candidate gene mapping via the QTL-seq approach (Zhao et al., 2020; Chen et al., 2021). An R2R3-MYB transcription factor gene named *AhTcl* was mapped and characterized as associated with purple testa via the QTL-seq approach. Allele-specific markers were developed, which demonstrated that the marker *pTesta1089* was closely linked with purple testa (Zhao et al., 2020). Chen et al. identified *AhRt1* bHLH transcriptional factor as the candidate gene that regulates the red testa color of peanut via QTL-sequencing analyses. An *AhRt1* diagnostic marker was then developed for validating and distinguishing different populations and peanut varieties (Chen et al., 2021). The phenotype evaluation of plant disease resistance traits is a challenge for map-based gene cloning. Unlike the testa color phenotype, disease resistance traits are complicated and affected by the environment, especially those for resistance to peanut bacterial wilt. As usual, the identification method in the disease nursery was used for the resistant evaluation. The survival rate was calculated from the number of dead plants until the point of harvesting for QTL mapping of BWR (Luo et al., 2019). However, the natural identification method was affected by the temperature, the soil environment, and anthropogenic effects. In our previous study, artificial inoculation by the leaf-cutting method was successfully used to evaluate the resistance to bacterial wilt via high throughput sequencing for QTL mapping of the cultivated peanut (Zhao et al., 2016). The resistance phenotyping in this study validated the accuracy of our previous findings (Zhao et al., 2016). The disease symptoms were classified into six disease severity ratings, and the resistance level of different lines was calculated by the disease index (Figure 1 and Supplementary Table 1). The phenotype of different lines was truly reflective of the disease resistance as per the DI method and the candidate genomic region was then accurately identified by the QTL-seq approach (Figure 2 and Supplementary Figures 6–9). This method is clearly valuable in the phenotyping of RRS for large populations in a cost-effective and practicable manner.

Hitherto, the main stable QTLs of RRS in the peanut were successfully identified through the original QTL method and QTL-seq approach (Luo et al., 2020). We previously reported SSR and SNP marker-based genetic linkage maps obtained through the classical QTL mapping (Zhao et al., 2016). Two major QTLs (*qBW-1* and *qBW-2*) were identified for RRS, which were in the LG1 and LG10 linkage groups based on RAD- and BSA-seq techniques in F_2 plants. One QTL linked to two QTL peaks on ChrB02 was identified in an F_8 RIL population (Zhao et al., 2016). Luo et al. reported one QTL named *qBWRB02.1* that possibly spanned a 5.14 Mb (0.81–5.95 Mb) interval on chromosome B02 based on its flanking SSR markers (Luo et al., 2020). Via the QTL-seq approach, they then identified a 2.07 Mb genomic region on ChrB02 associated with RRS across three environments (Luo et al., 2020). Two adjacent genomic regions (2.81–4.24 Mb and 6.54–8.75 Mb) on chromosome B02 were identified within the confidential interval of *qBWRB02-1* and

thus designated as *qBWRB02-1-1* and *qBWRB02-1-2* based on two diploids reference genomes (Luo et al., 2019). In the present study, by using the QTL-seq approach (Supplementary Figure 1), a major peak on Chr12 spanning a 7.2 Mb (1.8–9.0 Mb) interval with a confidence level of $P < 0.05$, 5.73 Mb of this peak had a confidence level of $P < 0.01$ was identified as the candidate genomic region for RRS, corroborating RAD-seq findings (Zhao et al., 2016) and SLAF-seq techniques (Figure 3 and Supplementary Figure 10, unpublished). These revealed the precise identification of the candidate genomic region via the QTL-seq approach. In China, peanut cultivars that are resistant to bacterial wilt, originate from Xiekangqing, Taishan Zhenzhu, and the wildtype species (*A. diogeni*). Two major QTLs that were both named *qBWRB02.1*, were identified from the cross of Yuanza 9102 \times Xuzhou 68-4. Yuanza 9102 is a popular resistant cultivar whose resistance stemmed from *A. diogeni* (Luo et al., 2020). Two major QTLs, *qBWRB02-1-1* and *qBWRB02-1-2*, were fine-mapped from the cross of Zhonghua 6 \times Xuhua 13. The resistance phenotype of Zhonghua 6 stemmed from the Chinese landrace Taishan Zhenzhu (Luo et al., 2019) whereas that of Yueyou92 stemmed from the Chinese landrace Xiekangqing, which is the parental type for many RRS breeding programs in South China (Janila et al., 2016; Luo et al., 2020). These candidate genomic regions associated with RRS were also mapped onto an interval of 10 Mb on chromosome 12.

QTL-seq approach was demonstrated as an effective method for the identification of putative SNPs associated with RRS. These could be developed into allele markers by using either different genotypes or diagnostic markers after the validation (Luo et al., 2019). Allele-specific markers that can be identified via agarose gel electrophoresis are the most cost-effective assays for genotyping a breeding population to select plants with the desired allele (Pandey et al., 2017). Here, 1807 effective SNPs and 629 InDels were identified. They span a 7.2 Mb genomic region on chromosome 12 that is associated with RRS. A total of 180 nonsynonymous SNPs and 14 InDels respectively affected 75 and 11 candidate genes that encode RRS (Tables 2, 3, Supplementary Tables 4 and 5). The putative RRS-encoding NBS-LRR gene had 44 SNPs, which were targeted for the development of allele-specific markers. Despite designing primers for both alleles of each SNP, amplification of markers was often observed for only one of the allele pairs. Nonamplification of a few markers may be due to DNA template-primer mismatches (You et al., 2008; Pandey et al., 2017). In this study, polymorphic SNP markers were selected as diagnostic markers. Of the 30 amplified markers, two markers (*qRRS18* and *qRRS19*) were robust and co-segregated with RRS (Figure 5). These two polymorphic markers were then validated on a panel of diverse genotypes containing naturally resistant parental types (Yueyou 92) of the RIL population, 18 introgression lines, susceptible parental types (Xinhuixiaoli), and 18 susceptible RIL lines. The '*qRRS18*' marker amplified susceptible alleles, whereas

the '*qRRS19*' marker amplified resistant alleles. Thus, these diagnostic allele markers can be applied in MAS for RRS in peanut breeding programs.

In plants, *R* genes play an important role in defending against pathogens through activation of the innate immune system (Jones and Dangl, 2006; Zhang et al., 2017; Sun et al., 2020; Chang et al., 2022). Most of them belong to the NBS-LRR type, which has been identified in many crops by map-based cloning (Takken and Joosten, 2000; Hulbert et al., 2001; McDowell and Woffenden, 2003; Meyers et al., 2005). The function of the NBS-LRR genes was correlated with either protein length or SNP variants. *RRS1-R* was the first reported Tir-NBS-LRR gene that conferred resistance to bacterial wilt in *Arabidopsis* species (Deslandes et al., 2003). *RRS1-S*, the allele of *RRS1-R* found in susceptible species, encode a protein without the WRKY domain. The two genes encoded contrasting phenotypes after *R. solanacearum* infections (Deslandes et al., 1998; Deslandes et al., 2003). Deng et al. identified an NBS-LRR gene named *PigmR*. It conferred resistance to the fungus *Magnaporthe oryzae* in rice and its encoded protein lacked four amino acids in the leucine-rich repeat (LRR) domain when compared to the *R4* gene that conferred a susceptible phenotype (Deng et al., 2017). In the present study, the predicted products of 22 of the 180 candidate genes with nonsynonymous mutations in the 7.2 Mb region, were NBS-LRR type disease resistance proteins (Table 2 and Figure 4). Notably, eight of the 22 candidate NBS-LRR genes were identified at a high confidence level as associated with RRS (Figure 4). Moreover, seven NBS-LRR genes had SNP variant sites in the LRR domain (Figure 4). The diagnostic SNP markers (*qRRS18* and *qRRS19*) of the candidate *AH12G03230* and *AH12G06320* NBS-LRR genes were validated in the RIL lines (Figure 5). This indicates that *AH12G03230* and *AH12G06320* might be the candidate resistant genes for RRS in cultivated peanut. Therefore, based on these findings, these putative resistance genes possibly significantly contribute to RRS in peanut and should thus be targeted as candidates for fine mapping and function validation.

5 Conclusion

In this study, the QTL-seq approach was proven as a powerful method for the successful identification of genomic regions and candidate genes of major and robust QTLs that are associated with RRS. We not only identified a 7.2 Mb genomic region on chromosome 12 containing eight candidate NBS-LRR resistance genes but also availed validated allele-specific diagnostic markers and key candidate genes for RRS breeding. The genomic information (genes) and tools (markers) could be used in genomics-assisted breeding programs to accelerate the development of peanut varieties with enhanced RRS as well as to increase insights into RRS molecular mechanisms.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA851221.

Author contributions

WZ and RV conceived the original research plan and designed the experiment. CZ, WX, HF, YC, HC, TC, QY, YZ, KC, and XZ performed the experiments and analyzed the data. CZ and WX wrote the manuscript, while WZ, MP, and RV reviewed and edit the manuscript. All authors analyzed the data and approved the submitted version.

Funding

This work was supported by National Science Foundation (NSF) of China (Grant No. 32072103 and 31701463 to CZ, Grant No. U1705233 to WZ), the Science and Technology Foundation of Fujian Province of China (Grant No. 2018N0004 to CZ), the Opening Grant of Guangdong Province Key Laboratory of Plant Molecular Breeding (Grant No. GPKLPMB202203 to CZ), the Special Fund for Scientific and Technological Innovation of Fujian Agriculture and Forestry University (Grant No. KFB22011XA to CZ).

References

- Arikiti, S., Wanchana, S., Khanthong, S., Saensuk, C., Thianthavon, T., Vanavichit, A., et al. (2019). QTL-seq identifies cooked grain elongation QTLs near soluble starch synthase and starch branching enzymes in rice (*Oryza sativa* L.). *Sci. Rep.* 9, 1–10. doi: 10.1038/s41598-019-44856-2
- Ben, C., Debellé, F., Berges, H., Bellec, A., Jardinaud, M.-F., Anson, P., et al. (2013). *MtQRRS1*, an R-locus required for *Medicago truncatula* quantitative resistance to *Ralstonia solanacearum*. *New Phytol.* 199, 758–772. doi: 10.1111/nph.12299
- Bommisetty, R., Chakravartty, N., Bodanapu, R., Naik, J. B., Panda, S. K., Lekkala, S. P., et al. (2020). Discovery of genomic regions and candidate genes for grain weight employing next generation sequencing based QTL-seq approach in rice (*Oryza sativa* L.). *mol. Biol. Rep.* 47, 8615–8627. doi: 10.1007/s11033-020-05904-7
- Cao, M., Li, S., Deng, Q., Wang, H., and Yang, R. (2021). Identification of a major-effect QTL associated with pre-harvest sprouting in cucumber (*Cucumis sativus* L.) using the QTL-seq method. *BMC Genomics* 22, 1–11. doi: 10.1186/s12864-021-07548-8
- Carmelle, A., Caranta, C., Dintinger, J., Prior, P., Luisetti, J., and Besse, P. (2006). Identification of QTLs for *Ralstonia solanacearum* race 3-phyto-type II resistance in tomato. *theor. Appl. Genet.* 113, 110–121. doi: 10.1007/s00122-006-0277-3
- Chang, M., Chen, H., Liu, F., and Fu, Z. Q. (2022). PTI and ETI: convergent pathways with diverse elicitors. *Trends Plant Sci.* 27, 113–115. doi: 10.1016/j.tplants.2021.11.013
- Chen, H., Chen, X., Xu, R., Liu, W., Liu, N., Huang, L., et al. (2021). Fine-mapping and gene candidate analysis for *AhRt1*, a major dominant locus responsible for testa color in cultivated peanut. *theor. Appl. Genet.* 134, 3721–3730. doi: 10.1007/s00122-021-03924-w
- Chen, Q., Song, J., Du, W. P., Xu, L. Y., Jiang, Y., Zhang, J., et al. (2018). Identification and genetic mapping for *rht-DM*, a dominant dwarfing gene in mutant semi-dwarf maize using QTL-seq approach. *Genes Genomics* 40, 1091–1099. doi: 10.1007/s13258-018-0716-y
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92. doi: 10.4161/fly.19695
- Clevenger, J., Chu, Y., Chavarro, C., Botton, S., Culbreath, A., Isleib, T. G., et al. (2018). Mapping late leaf spot resistance in peanut (*Arachis hypogaea*) using QTL-seq reveals markers for marker-assisted selection. *front. Plant Sci.* 9, 1–10. doi: 10.3389/fpls.2018.00083
- Das, S., Upadhyaya, H. D., Bajaj, D., Kujur, A., Badoni, S., Gowda, C. L. L., et al. (2014). 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
- Deng, Y., Zhai, K., Xie, Z., Yang, D., Zhu, X., Liu, J., et al. (2017). Epigenetic regulation of antagonistic receptors confers rice blast resistance with yield balance. *Science* 355, 962–965. doi: 10.1126/science.1238898
- Deslandes, L., Olivier, J., Peeters, N., Feng, D. X., Khounloham, M., Boucher, C., et al. (2003). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8024–8029. doi: 10.1073/pnas.1230660100

Acknowledgments

We would like to thank MogoEdit (<https://www.mogoeedit.com>) for its English editing during the preparation of this manuscript.

Conflict of interest

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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1048168/full#supplementary-material>

- Deslandes, L., Pileur, F., Liaubet, L., Camut, S., Beynon, J., Arlat, M., et al. (1998). "Identification and mapping of *RRS1*, a single recessive locus in *Arabidopsis thaliana* that confers resistance to *Ralstonia solanacearum*," in *Bacterial wilt disease* Springer, Berlin, Heidelberg, 250–254. doi: 10.1007/978-3-662-03592-4_36
- Dong, Z., Alam, M. K., Xie, M., Yang, L., Liu, J., Helal, M. M. U., et al. (2021). Mapping of a major QTL controlling plant height using a high-density genetic map and QTL-seq methods based on whole-genome resequencing in *Brassica napus*. *G3 (Bethesda)* 11, jkab118. doi: 10.1093/g3journal/jkab118
- Du, H., Wen, C., Zhang, X., Xu, X., Yang, J., Chen, B., et al. (2019). Identification of a major QTL (*qRRs-10.1*) that confers resistance to *Ralstonia solanacearum* in pepper (*Capsicum annuum*) using SLAF-BSA and QTL mapping. *Int. J. Mol. Sci.* 20, 5887. doi: 10.3390/ijms20235887
- Genin, S., and Boucher, C. (2004). Lessons learned from the genome analysis of *Ralstonia solanacearum*. *Annu. Rev. Phytopathol.* 42, 107–134. doi: 10.1146/annurev.phyto.42.011204.104301
- Godiard, L., Sauviac, L., Torii, K. U., Grenon, O., Mangin, B., Grimsley, N. H., et al. (2003). ERECTA, an LRR receptor-like kinase protein controlling development pleiotropically affects resistance to bacterial wilt. *Plant J.* 36, 353–365. doi: 10.1046/j.1365-3113X.2003.01877.x
- Habe, I., Miyatake, K., Nunome, T., Yamasaki, M., and Hayashi, T. (2019). QTL analysis of resistance to bacterial wilt caused by *Ralstonia solanacearum* in potato. *Breed. Sci.* 69, 592–600. doi: 10.1270/jsbbs.19059
- Hulbert, S. H., Webb, C. A., Smith, S. M., and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* 39, 285–312. doi: 10.1146/annurev.phyto.39.1.285
- Illa, E., Jason, B., and Huang, Z. (2015). Rapid and reliable identification of tomato fruit weight and locale number loci by QTL-seq. *Theor. Appl. Genet.*, 128, 1329–1342. doi: 10.1007/s00122-015-2509-x
- Janila, P., Variath, M. T., Pandey, M. K., Desmae, H., Motagi, B. N., Okori, P., et al. (2016). Genomic tools in groundnut breeding program: Status and perspectives. *Front. Plant Sci.* 7, 289. doi: 10.3389/fpls.2016.00289
- Jones, J. D. G., and Dangl, J. L. (2006). The plant immune system. *Nature* 444, 323–329. doi: 10.1038/nature05286
- Kumar, R., Janila, P., Vishwakarma, M. K., Khan, A. W., Manohar, S. S., Gangurde, S. S., et al. (2020). Whole-genome resequencing-based QTL-seq identified candidate genes and molecular markers for fresh seed dormancy in groundnut. *Plant Biotechnol. J.* 18, 992–1003. doi: 10.1111/pbi.13266
- Lebeau, A., Gouy, M., Daunay, M. C., Wicker, E., Chiroulet, F., Prior, P., et al. (2013). Genetic mapping of a major dominant gene for resistance to *Ralstonia solanacearum* in eggplant. *Theor. Appl. Genet.* 126, 143–158. doi: 10.1007/s00122-012-1969-5
- Lei, L., Zheng, H., Bi, Y., Yang, L., Liu, H., Wang, J., et al. (2020). Identification of a major QTL and candidate gene analysis of salt tolerance at the bud burst stage in rice (*Oryza sativa* L.) using QTL-seq and RNA-seq. *Rice* 13, 1–14. doi: 10.1186/s12284-020-00416-1
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- 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
- Luo, H., Pandey, M. K., Khan, A. W., Guo, J., Wu, B., Cai, Y., et al. (2019). Discovery of genomic regions and candidate genes controlling shelling percentage using QTL-seq approach in cultivated peanut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 17, 1248–1260. doi: 10.1111/pbi.13050
- Luo, H., Pandey, M. K., Khan, A. W., Wu, B., Guo, J., Ren, X., et al. (2019). Next-generation sequencing identified genomic region and diagnostic markers for resistance to bacterial wilt on chromosome B02 in peanut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 17, 2356–2369. doi: 10.1111/pbi.13153
- Luo, H., Pandey, M. K., Zhi, Y., Zhang, H., Xu, S., Guo, J., et al. (2020). Discovery of two novel and adjacent QTLs on chromosome B02 controlling resistance against bacterial wilt in peanut variety zhonghua 6. *Theor. Appl. Genet.* 133, 1133–1148. doi: 10.1007/s00122-020-03537-9
- 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
- Meyers, B. C., Kaushik, S., and Nandety, R. S. (2005). Evolving disease resistance genes. *Curr. Opin. Plant Biol.* 8, 129–134. doi: 10.1016/j.pbi.2005.01.002
- Mimura, Y., Kageyama, T., Minamiyama, Y., and Hirai, M. (2009). QTL analysis for resistance to *Ralstonia solanacearum* in *Capsicum* accession 'LS2341.' *J. Japanese Soc. Hort. Sci.* 78, 307–313. doi: 10.2503/jjshs1.78.307
- Narusaka, M., Hatakeyama, K., Shirasu, K., Narusaka, Y., Narusaka, M., Hatakeyama, K., et al. (2014). *Arabidopsis* dual resistance proteins, both RPS4 and RRS1, are required for resistance to bacterial wilt in transgenic Brassica crops. *Plant Signaling & Behav.* 9, e29130. doi: 10.4161/psb.29130
- Narusaka, M., Shirasu, K., Noutoshi, Y., Kubo, Y., Shiraiishi, T., Iwabuchi, M., et al. (2009). *RRS1* and *RPS4* provide a dual resistance-gene system against fungal and bacterial pathogens. *Plant J.* 60, 218–226. doi: 10.1111/j.1365-3113X.2009.03949.x
- Pandey, M. K., Khan, A. W., Singh, V. K., Vishwakarma, M. K., Shashidhar, Y., Kumar, V., et al. (2017). QTL-seq approach identified genomic regions and diagnostic markers for rust and late leaf spot resistance in groundnut (*Arachis hypogaea* L.). *Plant Biotechnol. J.* 15, 927–941. doi: 10.1111/pbi.12686
- Pandey, M. K., Pandey, A. K., Kumar, R., Nwosu, C. V., Guo, B., Wright, G. C., et al. (2020). Translational genomics for achieving higher genetic gains in groundnut. *Theor. Appl. Genet.* 133, 1679–1702. doi: 10.1007/s00122-020-03592-2
- Reddy, P. P. (2016). "Viral diseases and their management," in *Sustainable Crop Protection under Protected Cultivation* (Singapore: Springer press), 161–176.
- Reumers, J., De Rijk, P., Zhao, H., Liekens, A., Smeets, D., Cleary, J., et al. (2012). Optimized filtering reduces the error rate in detecting genomic variants by short-read sequencing. *Nat. Biotechnol.* 30, 61–68. doi: 10.1038/nbt.2053
- Salanoubat, M., Genin, S., Artiguenave, F., Gouzy, J., Manganot, S., Arlat, M., et al. (2002). Genome sequence *Plant pathogen Ralstonia solanacearum*. *Nature* 415, 497–502. doi: 10.1038/415497a
- Salgon, S., Jourda, C., Sauvage, C., Daunay, M. C., Reynaud, B., Wicker, E., et al. (2017). Eggplant resistance to the *Ralstonia solanacearum* species complex involves both broad-spectrum and strain-specific quantitative trait loci. *Front. Plant Sci.* 8, 828. doi: 10.3389/fpls.2017.00828
- Schell, M. A. (2000). Control of virulence and pathogenicity genes of *Ralstonia solanacearum* by an elaborate sensory network. *Annu. Rev. Phytopathol.* 38, 263–292. doi: 10.1146/annurev.phyto.38.1.263
- Shin, I. S., Hsu, J. C., Huang, S. M., Chen, J. R., Wang, J. F., Hanson, P., et al. (2020). Construction of a single nucleotide polymorphism marker based QTL map and validation of resistance loci to bacterial wilt caused by *Ralstonia solanacearum* species complex in tomato. *Euphytica* 216, 54. doi: 10.1007/s10681-020-2576-1
- 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
- Srivastava, R., Upadhyaya, H. D., Kumar, R., Daware, A., Basu, U., Shimray, P. W., et al. (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
- Sunkara, S., Bhatnagar-Mathur, P., and Sharma, K. K. (2014). "Transgenic interventions in peanut crop improvement: Progress and prospects," in *Genetics, genomics and breeding of crop plants* (Boca Raton, FL: CRC Press), 178–215.
- Sun, Y., Zhu, Y. X., Balint-Kurti, P. J., and Wang, G. F. (2020). Fine-tuning immunity: Players and regulators for plant NLRs. *Trends Plant Sci.* 25, 695–713. doi: 10.1016/j.tplants.2020.02.008
- Takagi, H., Abe, A., Yoshida, K., Kosugi, S., Natsume, S., Mitsuoka, 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
- Takken, F. L. W., and Joosten, M. H. A. J. (2000). Plant resistance genes: their structure, function and evolution. *Eur. J. Plant Pathol.* 106, 699–713. doi: 10.1023/A:1026571130477
- Thoquet, P., Olivier, J., Sperisen, C., Rogowsky, P., Laterrot, H., and Grimsley, N. (1996). Quantitative trait loci determining resistance to bacterial wilt in tomato cultivar Hawaii7996. *MPMI-Molecular Plant Microbe Interact.* 9, 826–836. doi: 10.1094/MPMI-9-0826
- Topcu, Y., Sapkota, M., Illa, E., Savithri, B., and Esther, U. N. (2021). Identification of blossom-end rot loci using joint QTL-seq and linkage-based QTL mapping in tomato. *Theor. Appl. Genet.* 134, 2931–2945. doi: 10.1007/s00122-021-03869-0
- Tudor, E. H., Jones, D. M., He, Z., Bancroft, I., Trick, M., Wells, R., et al. (2020). QTL-seq identifies *BnaFT.A02* and *BnaFLC.A02* as candidates for variation in vernalization requirement and response in winter oilseed rape (*Brassica napus*). *Plant Biotechnol. J.* 18, 2466–2481. doi: 10.1111/pbi.13421
- Varshney, R. K., Pandey, M. K., Bohra, A., Singh, V. K., Thudi, M., and Saxena, R. K. (2019). Toward the sequence-based breeding in legumes in the post-genome sequencing era. *Theor. Appl. Genet.* 132, 797–816. doi: 10.1007/s00122-018-3252-x
- Wang, H., Cheng, H., Wang, W., Liu, J., Hao, M., Mei, D., et al. (2016). Identification of *BnaYUCCA6* as a candidate gene for branch angle in *Brassica napus* by QTL-seq. *Sci. Rep.* 6, 38493. doi: 10.1038/srep38493
- Wang, J.-F., Ho, F.-I., Truong, H. T. H., Huang, S.-M., Balatero, C. H., Dittapongpich, V., et al. (2013). Identification of major QTLs associated with

stable resistance of tomato cultivar 'Hawaii 7996' to *Ralstonia solanacearum*. *Euphytica* 190, 241–252. doi: 10.1007/s10681-012-0830-x

Wang, Y., Sun, H., Wang, H., Yang, X., Xu, Y., Yang, Z., et al. (2021). Integrating transcriptome, co-expression and QTL-seq analysis reveals that primary root growth in maize is regulated via flavonoid biosynthesis and auxin signal transduction. *J. Exp. Bot.* 72, 4773–4795. doi: 10.1093/jxb/erab177

Yang, L., Wang, J., Han, Z., Lei, L., Liu, H. L., Zheng, H., et al. (2021). Combining QTL-seq and linkage mapping to fine map a candidate gene in *qCTS6* for cold tolerance at the seedling stage in rice. *BMC Plant Biol.* 21, 1–14. doi: 10.1186/s12870-021-03076-5

You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D., et al. (2008). BatchPrimer3: A high throughput web application for PCR and sequencing primer design. *BMC Bioinf.* 9, 253. doi: 10.1186/1471-2105-9-253

Yu, S. L., Wang, C. T., Yang, Q. L., Zhang, D. X., Zhang, X. Y., Cao, Y. L., et al. (2011). *Peanut genetics and breeding in China* Vol. 565 (Shanghai: Shanghai Sci. Technol. Press).

Zhang, C., Badri Anarjan, M., Win, K. T., Begum, S., and Lee, S. (2021). QTL-seq analysis of powdery mildew resistance in a Korean cucumber inbred line. *Theor. Appl. Genet.* 134, 435–451. doi: 10.1007/s00122-020-03705-x

Zhang, C., Chen, H., Cai, T., Deng, Y., Zhuang, R., Zhang, N., et al. (2017). Overexpression of a novel peanut NBS-LRR gene *AhRRS5* enhances disease resistance to *Ralstonia solanacearum* in tobacco. *Plant Biotechnol. J.* 15, 39–55. doi: 10.1111/pbi.12589

Zhang, C., Chen, H., Zhuang, R. R., Chen, Y. T., Deng, Y., Cai, T. C., et al. (2019). Overexpression of the peanut CLAVATA1-like leucine-rich repeat receptor-like kinase *AhRLK1* confers increased resistance to bacterial wilt in tobacco. *J. Exp. Bot.* 70, 5407–5421. doi: 10.1093/jxb/erz274

Zhao, Y., Ma, J., Li, M., Deng, L., Li, G., Xia, H., et al. (2020). Whole-genome resequencing-based QTL-seq identified *AhTc1* gene encoding a R2R3-MYB transcription factor controlling peanut purple testa colour. *Plant Biotechnol. J.* 18, 96–105. doi: 10.1111/pbi.13175

Zhao, Y., Zhang, C., Chen, H., Yuan, M., Nipper, R., Prakash, C. S., et al. (2016). QTL mapping for bacterial wilt resistance in peanut (*Arachis hypogaea* L.). *Mol. Breed* 36, 1–11. doi: 10.1007/s11032-015-0432-0

Zhuang, W., Chen, H., Yang, M., Wang, J., Pandey, M. K., Zhang, C., et al. (2019). The genome of cultivated peanut provides insight into legume karyotypes, polyploid evolution and crop domestication. *Nat. Genet.* 51, 865–876. doi: 10.1038/s41588-019-0402-2

Frontiers in Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

[See more →](#)

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

