

NOVEL TECHNOLOGIES FOR SOYBEAN IMPROVEMENT

EDITED BY: Xianzhong Feng, Deyue Yu and Madan K. Bhattacharyya
PUBLISHED IN: Frontiers in Plant Science





frontiers

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 978-2-83250-399-7

DOI 10.3389/978-2-83250-399-7

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

NOVEL TECHNOLOGIES FOR SOYBEAN IMPROVEMENT

Topic Editors:

Xianzhong Feng, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences (CAS), China

Deyue Yu, Nanjing Agricultural University, China

Madan K. Bhattacharyya, Iowa State University, United States

Citation: Feng, X., Yu, D., Bhattacharyya, M. K., eds. (2022). Novel Technologies for Soybean Improvement. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-399-7

Table of Contents

- 05 Editorial: Novel Technologies for Soybean Improvement**
Xianzhong Feng, Deyue Yu and Madan K. Bhattacharyya
- 09 Fine-Tuning Florigen Increases Field Yield Through Improving Photosynthesis in Soybean**
Kun Xu, Xiao-Mei Zhang, Haifeng Chen, Chanjuan Zhang, Jinlong Zhu, Zhiyuan Cheng, Penghui Huang, Xinan Zhou, Yuchen Miao, Xianzhong Feng and Yong-Fu Fu
- 22 Integrated Transcriptomic and Bioinformatics Analyses Reveal the Molecular Mechanisms for the Differences in Seed Oil and Starch Content Between Glycine max and Cicer arietinum**
Kun Cheng, Yi-Fan Pan, Lü-Meng Liu, Han-Qing Zhang and Yuan-Ming Zhang
- 39 Optimizing RNAi-Target by Nicotiana benthamiana-Soybean Mosaic Virus System Drives Broad Resistance to Soybean Mosaic Virus in Soybean**
Hua Jiang, Kai Li and Junyi Gai
- 52 Mutation of GmAIR Genes by CRISPR/Cas9 Genome Editing Results in Enhanced Salinity Stress Tolerance in Soybean**
Tianya Wang, Hongwei Xun, Wei Wang, Xiaoyang Ding, Hainan Tian, Saddam Hussain, Qianli Dong, Yingying Li, Yuxin Cheng, Chen Wang, Rao Lin, Guimin Li, Xueyan Qian, Jinsong Pang, Xianzhong Feng, Yingshan Dong, Bao Liu and Shucai Wang
- 67 Identification of Finely Mapped Quantitative Trait Locus and Candidate Gene Mining for the Three-Seeded Pod Trait in Soybean**
Candong Li, Hongwei Jiang, Yingying Li, Chunyan Liu, Zhaoming Qi, Xiaoxia Wu, Zhanguo Zhang, Zhenbang Hu, Rongsheng Zhu, Tai Guo, Zhixin Wang, Wei Zheng, Zhenyu Zhang, Haihong Zhao, Nannan Wang, Dapeng Shan, Dawei Xin, Feishi Luan and Qingshan Chen
- 81 Establishment and Application of Ligation Reaction-Based Method for Quantifying MicroR-156b**
Yuxuan He, Likun Long, Wei Yan, Liming Dong, Wei Xia, Congcong Li and Feiwu Li
- 92 WinRoots: A High-Throughput Cultivation and Phenotyping System for Plant Phenomics Studies Under Soil Stress**
Yangyang Zhang, Wenjing Zhang, Qicong Cao, Xiaojian Zheng, Jingting Yang, Tong Xue, Wenhao Sun, Xinrui Du, Lili Wang, Jing Wang, Fengying Zhao, Fengning Xiang and Shuo Li
- 105 Deciphering Novel Transcriptional Regulators of Soybean Hypocotyl Elongation Based on Gene Co-expression Network Analysis**
Zhikang Shen and Min Chen
- 118 CRISPR/Cas9-Mediated Targeted Mutagenesis of GmUGT Enhanced Soybean Resistance Against Leaf-Chewing Insects Through Flavonoids Biosynthesis**
Yongxing Zhang, Wei Guo, Limiao Chen, Xinjie Shen, Hongli Yang, Yisheng Fang, Wenqi Ouyang, Sihua Mai, Haifeng Chen, Shuilian Chen, Qingnan Hao, Songli Yuan, Chanjuan Zhang, Yi Huang, Zhihui Shan, Zhonglu Yang, Dezhen Qiu, Xinan Zhou, Dong Cao, Xia Li and Yongqing Jiao

- 132 ***Traits Expansion and Storage of Soybean Phenotypic Data in Computer Vision-Based Test***
Yongchao Xing, Peixin Lv, Hong He, Jiantian Leng, Hui Yu and Xianzhong Feng
- 149 ***Rapid Identification of Soybean Varieties by Terahertz Frequency-Domain Spectroscopy and Grey Wolf Optimizer-Support Vector Machine***
Xiao Wei, Dandan Kong, Shiping Zhu, Song Li, Shengling Zhou and Weiji Wu
- 158 ***Genomic Prediction of Green Fraction Dynamics in Soybean Using Unmanned Aerial Vehicles Observations***
Yusuke Toda, Goshi Sasaki, Yoshihiro Ohmori, Yuji Yamasaki, Hirokazu Takahashi, Hideki Takanashi, Mai Tsuda, Hiromi Kajiya-Kanegae, Raul Lopez-Lozano, Hisashi Tsujimoto, Akito Kaga, Mikio Nakazono, Toru Fujiwara, Frederic Baret and Hiroyoshi Iwata
- 170 ***Full-Length Transcriptional Analysis of the Same Soybean Genotype With Compatible and Incompatible Reactions to *Heterodera glycines* Reveals Nematode Infection Activating Plant Defense Response***
Minghui Huang, Ye Jiang, Ruifeng Qin, Dan Jiang, Doudou Chang, Zhongyan Tian, Chunjie Li and Congli Wang
- 193 ***Soybean F-Box-Like Protein GmFBL144 Interacts With Small Heat Shock Protein and Negatively Regulates Plant Drought Stress Tolerance***
Keheng Xu, Yu Zhao, Yan Zhao, Chen Feng, Yinhe Zhang, Fawei Wang, Xiaowei Li, Hongtao Gao, Weican Liu, Yan Jing, Rachit K. Saxena, Xianzhong Feng, Yonggang Zhou and Haiyan Li
- 206 ***Progress in Soybean Genetic Transformation Over the Last Decade***
Hu Xu, Yong Guo, Lijuan Qiu and Yidong Ran
- 239 ***Identification of Candidate Genes and Genomic Selection for Seed Protein in Soybean Breeding Pipeline***
Jun Qin, Fengmin Wang, Qingsong Zhao, Ainong Shi, Tiantian Zhao, Qijian Song, Waltram Ravelombola, Hongzhou An, Long Yan, Chunyan Yang and Mengchen Zhang
- 251 ***Analysis of Physiological Variations and Genetic Architecture for Photosynthetic Capacity of Japanese Soybean Germplasm***
Mohammad Jan Shamim, Akito Kaga, Yu Tanaka, Hiroshi Yamatani and Tatsuhiko Shiraiwa
- 264 ***Improving Ultra-Low Temperature Preservation Technologies of Soybean Pollen for Off-Season and Off-Site Hybridization***
Hongchang Jia, Xin Liang, Lixin Zhang, Jinmei Zhang, Enoch Sapey, Xianyuan Liu, Yanhui Sun, Shi Sun, Hongrui Yan, Wencheng Lu and Tianfu Han
- 276 ***Artificial Neural Network for Discrimination and Classification of Tropical Soybean Genotypes of Different Relative Maturity Groups***
Lígia de Oliveira Amaral, Glauco Vieira Miranda, Bruno Henrique Pedroso Val, Alice Pereira Silva, Alyce Carla Rodrigues Moitinho and Sandra Helena Unêda-Trevisoli
- 284 ***Exploring Soybean Flower and Pod Variation Patterns During Reproductive Period Based on Fusion Deep Learning***
Rongsheng Zhu, Xueying Wang, Zhuangzhuang Yan, Yinglin Qiao, Huilin Tian, Zhenbang Hu, Zhanguo Zhang, Yang Li, Hongjie Zhao, Dawei Xin and Qingshan Chen



OPEN ACCESS

EDITED AND REVIEWED BY
Marcos Egea-Cortines,
Universidad Politécnica de Cartagena,
Spain

*CORRESPONDENCE
Xianzhong Feng
fengxianzhong@iga.ac.cn

SPECIALTY SECTION
This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

RECEIVED 28 September 2022

ACCEPTED 03 October 2022

PUBLISHED 17 October 2022

CITATION
Feng X, Yu D and Bhattacharyya MK
(2022) Editorial: Novel technologies
for soybean improvement.
Front. Plant Sci. 13:1047739.
doi: 10.3389/fpls.2022.1047739

COPYRIGHT
© 2022 Feng, Yu and Bhattacharyya.
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: Novel technologies for soybean improvement

Xianzhong Feng^{1*}, Deyue Yu² and Madan K. Bhattacharyya³

¹Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China, ²National Key Laboratory of Crop Genetics and Germplasm Enhancement, National Center for Soybean Improvement, Nanjing Agricultural University, Nanjing, China, ³Department of Agronomy, Iowa State University, Ames, IA, United States

KEYWORDS

editorial, novel technologies, for, soybean, improvement

Editorial on the Research Topic

Novel technologies for soybean improvement

Introduction

Growing human population has put enormous pressure on the global food security. In the past few decades, the soybean yield has remained stagnant especially due to the use of conventional breeding technologies. In this regard, the research community were looking for the novel breeding technologies to revolutionize the soybean breeding. “Novel breeding technologies” are biotech-based approaches to modify plant characteristics fast and accurately. In this scenario, the current Research Topic “novel technologies for soybean improvement” is intended to collect articles on recent advances and future applications of the novel technologies in the soybean breeding to boost the soybean production. These approaches and technologies include molecular design breeding approaches, genome editing and transformation technology, RNA interference approach, Marker-Assisted and Genomics-Selection breeding approaches, machine learning and bioinformatics technology. Below we briefly highlight the applications and potential of the new approaches in the soybean breeding; and how they can be significant for increasing the yield and quality in soybean that are presented in a collection of 20 papers published in the special issue on the Research Topic: Novel Technologies for Soybean Improvement.

Computation technology has been put to use widely for soybean improvement

Soybean flower and pod drop are important determinants of soybean yield, thus use of advanced techniques that will increase the accuracy and speed for the flower and pod phenotyping has great influence on soybean breeding. In this regard, the computer vision techniques have recently emerged to phenotyping the flowers and pods in bulk at higher speed and accuracy. [Zhu et al. \(2022\)](#) identified that among the various deep learning algorithms, the

Faster R-CNN model performance was the best to phenotype soybean flowers and pods. The accuracy of Faster R-CNN model was 94.36 and 91% for detecting the flowers and pods, respectively. Furthermore, based on the Faster R-CNN model, they also proposed the fusion model for soybean flower and pod recognition and counting, that will greatly reduce labor intensity and improve efficiency.

Identification of the soybean varieties with superior nutritional value and composition is an important goal of soybean researchers. In this regard, [Wei et al. \(2022\)](#) checked the performance of Terahertz frequency-domain spectroscopy and chemometrics for the identification of soybean varieties, and subsequently proposed the grey wolf optimizer-support vector machine (GWO-SVM) soybean variety identification model. These authors showed that combination of the Terahertz frequency-domain spectroscopy and GWO-SVM will identify the soybean varieties at higher accuracy and speed. Compared with discriminant partial least squares (DPLS) and particles swarm optimization support vector machine, GWO-SVM combined with the second derivative could establish a better soybean variety identification model.

Soybean possessing narrow genetic base makes it hard to screen the available genetic and phenotypic variability for identifying the superior genotypes. In this regard, “Artificial Neural Network” (ANN) is the type of machine learning process which is used to classify the narrow-range and pure lineage populations. [Amaral et al. \(2022\)](#) argue that ANN can be used to categorize soybean genotypes from a population that has an either a wide genetic variability or a narrow genetic variability for the relative maturity character. The authors further demonstrated that ANN can be used to discriminate genotypes in early breeding generations. This would make it easier and to rapidly find high-performing cultivars with longer and shorter photoperiod amplitudes at a single selection site.

Artificial intelligence (AI) technology has been expected to provide the accurate phenotypic data at high resolution and low cost in the plants. [Xing et al. \(2022\)](#) used the computer vision (CV), machine learning based technology for the phenotypic analysis in soybean. They focused on expanding four different kinds of features disease traits, indoor test traits, field traits and soybean seed phenotypic traits. It was revealed that CV technology has the ability to collect highly accurate characteristics, and allows to provide large quantity of big data for breeding programs.

More and more novel technologies of both phenomics, genomics and bioinformatics have been applied in soybean improvement

Availability of high-throughput phenotyping systems has provided the opportunity to have an easy access to growth data.

This will in turn allow to predict growth of cultivars with unknown performance by conducting the genomic prediction analysis. Till now, the genomic and growth modeling prediction have not been studied to the greater depth. In this regard, [Toda et al. \(2022\)](#) used the data collected by an Unmanned Aerial Vehicle (UAV) to predict genomic green fraction dynamics in soybean. In their study, the use of UAV remote sensing allowed the measurement of the longitudinal variations in the green fractions. Their study showed that the model effectively combined early growth data with training population phenotypic data for prediction. This prediction method could be applied to selection at an early growth stage in crop breeding, and could reduce the cost and time of field trials.

Salinity is one of the major environmental constraints effecting the crop yield. However, there is considerable lack of high throughput phenomics platforms that allow to record the salt stress responses of plants as well as the non-destructive collection method for root phenotypic data. [Zhang et al. \(2022\)](#) reported the high-throughput and low-cost phenotypic platform that allows phenotyping whole plant including roots at uniform and controlled soil stress conditions. They concluded that responses of cotyledons can be used as non-destructive indicator for determining the salt tolerance of the seedlings based on their high-throughput multiple-phenotypic assays of 178 soybean cultivars.

Photosynthesis is an important process that determines the plant growth and yield. Hence, it is prerequisite to more efficiently assay the photosynthetic capacity of crop plants in order to select the varieties with higher photosynthetic efficiency. [Shamim et al. \(2022\)](#) explored the underlying physiological mechanisms and genetic basis regulating the photosynthesis in the Japanese soybean germplasm by using the “MIC-100” system (high-throughput system for measuring the photosynthesis). They identified the signification association of single nucleotide polymorphism (SNPs) on Chromosome 17 with the light-saturated photosynthesis (Asat) trait based on genome wide association study (GWAS) analysis. The G protein alpha subunit 1 (GPA1) was identified as the strong candidate Asat gene. This study provides strong evidence to apply GWAS of plant germplasm for exploring the maximum potential of photosynthesis in soybean improvement.

Limited studies have been conducted to determine the molecular mechanism underlying the seed oil and starch content in soybean. These traits are important traits determining the yield and quality of soybean. [Cheng et al. \(2021\)](#) used an integrated transcriptomic approach for investigating the species-specific, starch-related, carbon metabolism-related and acyl-lipid-related genes in soybean and chickpea. They identified seven soybean-specific gene expression patterns, four of which are highly expressed at the middle- and late oil accumulation stages. They proposed the difference of metabolism pathways in seed oil and starch content between soybean and chickpea, this study has opened up the

possibility of engineering other legumes for enhanced oil contents.

Transcriptome sequencing of full-length transcripts by long sequencing read method is an efficient approach to reveal detailed information about the events at transcriptional or post-transcriptional levels. But this approach has not been used in the soybean till now. In this context, [Huang et al. \(2022\)](#) used this approach to compare full-length transcriptomes in soybean genotype 09-138 infected with either soybean cyst nematode (*Heterodera glycines*) race 4 or race 5 that are avirulent and virulent, respectively to the soybean genotype. This study provides the insights about the soybean-nematode interaction which will serve as the basis for future research.

Seed germination is regulated by the hypocotyl elongation, and it determines the vitality of seedlings; however, the mechanism underlying the hypocotyl elongation has remained largely unexplored. [Shen and Chen \(2022\)](#) used the weighted gene co-expression network analysis (WGCNA) for the first time to understand the global regulatory network of gene expression involved in soybean hypocotyl elongation and identified a crucial regulatory module. They reported two regulatory modules. GmPRE6s-EXPANSINs submodule and GmPIF1/GmPIF3-GmSAUR1/23-EXPANSINs submodule. This study provides important genes for developing molecular markers for breeding for soybean with enhanced hypocotyl length and seed vigor.

Role of the miRNAs in regulating biological processes as well as biotic and abiotic stress responses in plants has been well documented. Hence, identification of the miRNAs is essential to determine their role in the biological processes as well as for plant responses to environmental stresses in plants. [He et al. \(2021\)](#) developed a highly sensitive approach for quantitatively detecting miRNAs by ligating ribonucleotide-modified DNA probes. This approach eliminates the use of complex reverse transcription-based quantitative approach. In the devised test, miR156b immediately hybridized two ribonucleotide-modified DNA probes, and amplification using universal primers followed the ligation process. The target miRNA could be detected at a 0.0001 amol level, and variations of a single base between miR156 family members could be recognized. The suggested quantitative technique works for overexpression-based genetically modified (GM) soybean. Ligation-based quantitative polymerase chain reaction (PCR) may be used to study miRNAs and investigate GM organisms.

In plants, the F-box gene family is one of the big gene families. F-box genes are involved in the regulation of plant growth as well as response to multiple environmental stresses, both biotic and abiotic. [Xu et al. \(2022\)](#) identified the F-box gene family in soybean genome using the bioinformatic approach, and also elucidated the role of F-box-like gene *GmFBL144* in adapting soybean to drought stress. They identified 507 F-box genes that were grouped into 11 subfamilies. The *GmFBL144*

showed higher expression in the roots and ectopic expression of the *GmFBL144* in *Arabidopsis* enhanced the sensitivity to drought stress. Their study suggested that *GmFBL144* protein negatively regulates drought stress through its interaction with small heat shock protein.

Approaches of genome-wide analysis are becoming the mainstream for soybean gene function study

To breed soybean varieties with high seed protein, it is a prerequisite to elucidate the genetic basis of high seed protein. [Qin et al. \(2022\)](#) used the combined approach of linkage mapping and genome-wide association studies (GWAS) to identify the QTLs and genes underlying the seed protein in soybean. They identified major QTLs mapped to Chromosomes 6 and 20 in two segregating populations. Their study revealed the similar performance of both BL and rrBLUP, and the accuracy of GS was dependent on size of training population and SNP set. The protein-specific SNPs showed higher GS efficiency compared to that by all SNPs.

The three-seeded pod number is an important trait that positively influences soybean yield. Soybean variety with increased three-seeded pod number contributes to the seed number/plant and higher yield. The candidate genes of the three-seeded pod may be the key for improving soybean yield. In this regard, [Li et al. \(2021\)](#) identified and validated the candidate genes regulating the three-seeded pod. They used the QTL mapping approach and identified 36 QTL. Four genes from 162 genes underlying these two QTL related to pod and seed set were identified as potential candidate genes controlling this trait. These four genes can be used as molecular markers for incorporating this trait into commercial soybean cultivars.

Genome editing and transformation technology are reaching to maturity to reap benefits for soybean improvement

In the past four decades the genetic transformation was used to modify the genome of soybean for the trait improvement. However, the genetic transformation in soybean is still subjected to low transformation efficiency, genotype specificity as well as long and tedious transformation processes. [Xu et al. \(2022\)](#) reviewed the genetic transformation progress in soybean and reported the gain made in efficiency and genotype flexibility of soybean transformation over the last decade. Factors that influence the soybean transformation are proper explant

selection, reagent selection, components of culture medium and the detailed understanding of the mechanism involved in the transformation.

CRISPR/Cas9 genome editing has emerged an efficient technique to breeding crop cultivars with enhance tolerance to environmental stresses. In Arabidopsis, knockout of the entire family of the abscisic acid (ABA)-induced transcription repressors (AITRs) enhanced drought and salinity tolerance without any fitness costs. Wang et al. (2021) demonstrated that mutation of *GmAITS* genes by CRISPR/Cas9 genome editing in soybean increased the salinity tolerance. The enhanced salt tolerance was observed both at the seed germination and seedlings stages of *gmaitr* mutants. Hence, the knockout of the *GmAITS* genes via CRISPR/Cas9 proved to be a powerful approach for inducing salt tolerance in soybean.

Insect pests such as leaf-chewing insects are one of the major constraints in soybean production. In this context, the development of soybean varieties with increased resistance to leaf-chewing insects can greatly increase the soybean yield. The *Glyma.07g110300* (*GmUGT*) encodes an UDP-glycosyltransferase (UGT) which is the core enzyme that negatively regulates insect resistance. Zhang et al. (2022) demonstrated that CRISPR/Cas9-mediated mutagenesis/knockout of *GmUGT* enhanced soybean resistance to *Helicoverpa armigera* and *Spodoptera litura*. The ectopic expression of the *GmUGT* gene in Arabidopsis *ugt72b1* mutant this mutant significantly reduced its resistance to *H. armigera*.

Soybean mosaic virus (SMV) is a prevalent soybean pathogen. Pyramiding multiple SMV-resistance genes into one individual is tedious and difficult. Targeting the viral genome via host-induced gene silencing (HIGS) has potential to induce broad-spectrum resistance (BSR) to SMV. Jiang et al. (2021) employed the *Nicotiana benthamiana*-soybean mosaic virus (SMV) system to optimize the target SMV sequence for HIGS. They were able to use the information gained from the *N. benthamiana*-SMV system to develop transgenic soybean lines with effective HIGS to enhance SMV resistance in soybean.

Florigen is a key player regulating the balance between vegetative and reproductive growth. Hence, it regulates the crop yield; but the actual utilization of the *florigen* in improving the crop yield remains largely unexplored. In this context, Xu et al. (2021) developed a strategy to improve the yield of soybean significantly by using the *florigen*. They showed that fine-tuning of the *florigen* genes via RNAi can boosts

soybean yield. The conserved functions of *florigen* in plants and use of the RNAi approach to manipulate the expression of this gene therefore may be used to increase crop productivity.

To maintain the crop germplasm for its use in future breeding programs, preservation of the pollen grains via ultra-low temperature approach is an effective and safe way for long-term storage of plant germplasm resources. Jia et al. (2022) reported an efficient ultra-low temperature freezing approach to preserve soybean pollens. Soybean flowers collected at the fully-bloom stage were dried, frozen and stored at -196 or -80°C. They observed that 90% of the pollen grains remain viable based on *in vitro* tests. Their study documented that this technique of pollen preservation will break the spatiotemporal barrier of soybean hybridization and facilitate efficient soybean breeding.

Author contributions

XF wrote the draft, DY and MB edited. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by National Nature Science Foundation of China (No. U21A20215).

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.



Fine-Tuning Florigen Increases Field Yield Through Improving Photosynthesis in Soybean

Kun Xu^{1,2†}, Xiao-Mei Zhang^{1†}, Haifeng Chen^{3†}, Chanjuan Zhang³, Jinlong Zhu^{1,2}, Zhiyuan Cheng¹, Penghui Huang¹, Xinan Zhou³, Yuchen Miao⁴, Xianzhong Feng^{5*} and Yong-Fu Fu^{1*}

¹ MOA Key Laboratory of Soybean Biology, National Key Facility of Crop Gene Resource and Genetic Improvement, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ² Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Harbin, China, ³ Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture, Wuhan, China, ⁴ Key Laboratory of Plant Stress Biology, State Key Laboratory of Cotton Biology, School of Life Sciences, Henan University, Kaifeng, China, ⁵ CAS Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

OPEN ACCESS

Edited by:

Scott A. Boden,
John Innes Centre, United Kingdom

Reviewed by:

Kyuya Harada,
Osaka University, Japan
Yaroslav B. Blume,
National Academy of Sciences
of Ukraine (NAN Ukraine), Ukraine

*Correspondence:

Xianzhong Feng
fengxianzhong@iga.ac.cn
Yong-Fu Fu
fuyongfu@caas.cn

[†] These authors have contributed
equally to this work

Specialty section:

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

Received: 17 May 2021

Accepted: 28 June 2021

Published: 16 August 2021

Citation:

Xu K, Zhang X-M, Chen H,
Zhang C, Zhu J, Cheng Z, Huang P,
Zhou X, Miao Y, Feng X and Fu Y-F
(2021) Fine-Tuning Florigen Increases
Field Yield Through Improving
Photosynthesis in Soybean.
Front. Plant Sci. 12:710754.
doi: 10.3389/fpls.2021.710754

Crop yield has been maintaining its attraction for researchers because of the demand of global population growth. Mutation of flowering activators, such as florigen, increases plant biomass at the expense of later flowering, which prevents crop maturity in the field. As a result, it is difficult to apply flowering activators in agriculture production. Here, we developed a strategy to utilize florigen to significantly improve soybean yield in the field. Through the screening of transgenic lines of RNAi-silenced florigen homologs in soybean (*Glycine-max-Flowering Locus T Like*, *GmFTL*), we identified a line, *GmFTL*-RNAi#1, with minor changes in both *GmFTL* expression and flowering time but with notable increase in soybean yield. As expected, *GmFTL*-RNAi#1 matured normally in the field and exhibited markedly high yield over multiple locations and years, indicating that it is possible to reach a trade-off between flowering time and high yield through the fine-tuning expression of flowering activators. Further studies uncovered an unknown mechanism by which *GmFTL* negatively regulates photosynthesis, a substantial source of crop yield, demonstrating a novel function of florigen. Thus, because of the highly conserved functions of florigen in plants and the classical RNAi approach, the findings provide a promising strategy to harness early flowering genes to improve crop yield.

Keywords: high yield, florigen, FT, photosynthesis, soybean, vegetative growth

INTRODUCTION

The global crop demand for human consumption and livestock feed is forecasted to increase by 110% from 2005 to 2050 (Tilman et al., 2011). However, advances in breeding, genomics, and transgenic technology are predicted to increase yield by up to 20% (Long et al., 2015). Therefore, there is an urgent need to develop innovative approaches to increase crop yields. Photosynthesis provides a substantial means to adjust crop yields (Ort et al., 2015). Regulation of photosynthesis happens at multiple levels (Long et al., 2006; Parry et al., 2007), which could serve as targets to

enhance the efficiency of photosynthesis and, therefore, crop yields (Ort et al., 2015; South et al., 2019). Increasing photosynthetic efficiency would include improving the photosynthetic process through changing the structure and physiology of the chloroplast in multiple targets, therefore avoiding potentially harmful effects from the alteration of a single factor.

Unsurprisingly, flowering time is widely used as a selectable marker in high-yield plant breeding programs (Blumel et al., 2015). A long vegetative phase means later flowering and high yield because it provides numerous resources for increased yield. However, crops in the field may not mature normally before winter if the vegetative phase is too long. Therefore, balancing vegetative growth and reproductive growth will achieve high yield in a normal growth season. The transition from the vegetative to the reproductive phase is regulated by a complex genetic network. Plant monitors and integrates both the developmental and environmental signals to produce florigen (*Flowering Locus T*, *FT*) (Yoo et al., 2005; Corbesier et al., 2007; Tamaki et al., 2007; Andres and Coupland, 2012). The lower the florigen production, the later the flowering and the higher the yield (Andres and Coupland, 2012; Blumel et al., 2015; Cho et al., 2017). The *FT* dosage plays a key role in the yield of tomato (Krieger et al., 2010) and rice (Huang et al., 2016). However, utilizing flowering genes, such as the *FT* gene, to increase crop yield in the field remains unknown.

In addition to flowering control, *FT* homologs also contribute to the regulation of vegetative growth, namely, tuberization (Navarro et al., 2011), onion bulb formation (Lee et al., 2013), sugar beet growth (Pin et al., 2010), and seed dormancy (Chen and Penfield, 2018). However, the understanding of *FT* in regulating leaf growth is limited, even though it is reported that overexpression of *FT* leads to smaller leaves and reduced expression leads to increased leaf size in *Arabidopsis* (Teper-Bamnolker and Samach, 2005). There is no report showing the relationship between the *FT* gene and photosynthesis.

Here, we show that a decreased expression of the florigen gene significantly enhances chloroplast and leaf development, as well as photosynthetic efficiency, and therefore increases soybean yield. We further test the effect of silencing florigen on field soybean yield, and screen out the transgenic line *GmFTL*-RNAi line #1, which shows flowering time quite similar to that of wild-type plants and has significantly higher yield in the field. Therefore, we develop a smart strategy to utilize flowering activators for crop yield improvement.

RESULTS

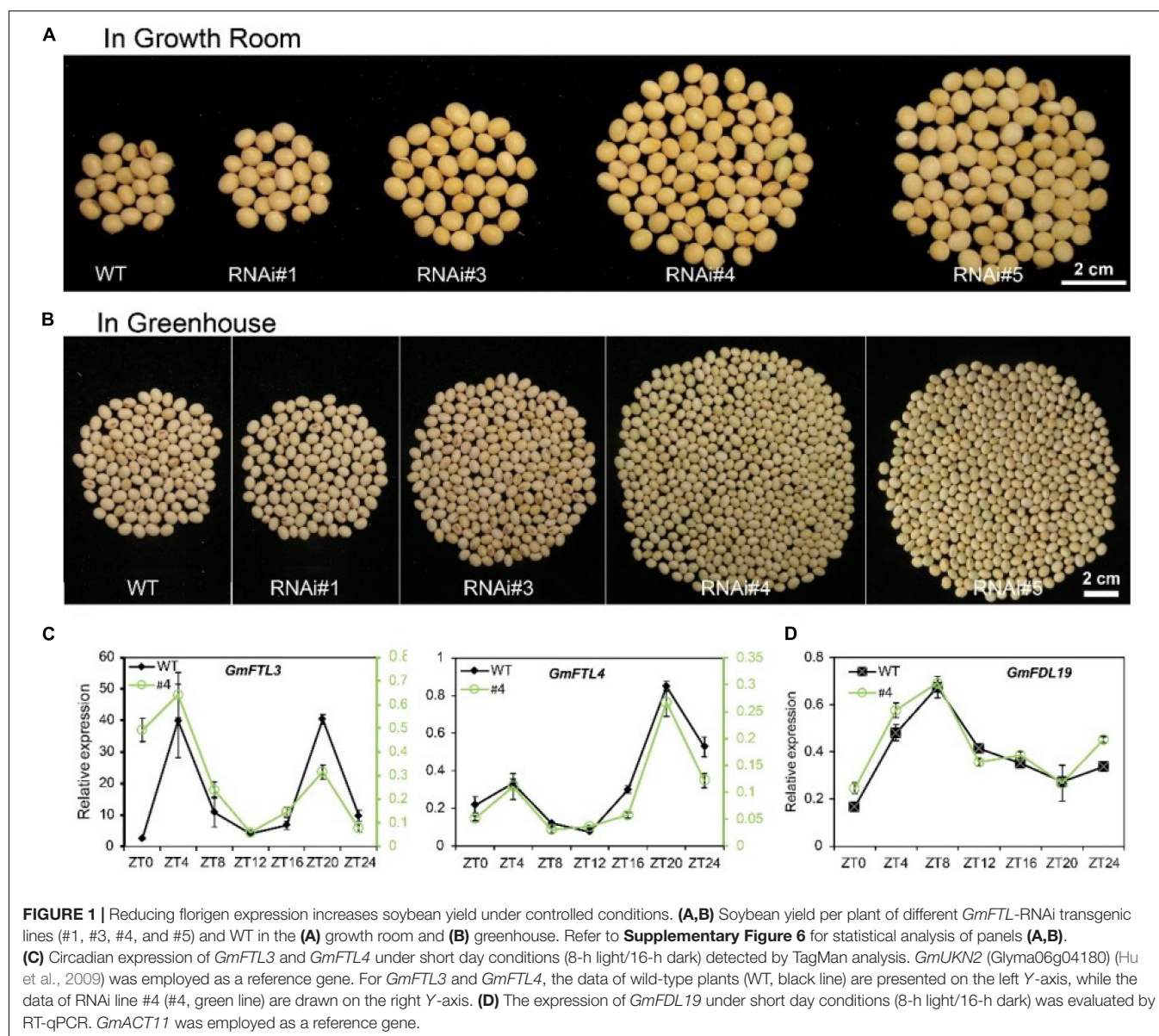
GmFTL Expression Is Negatively Correlated With Soybean Yield Under Controlled Conditions

GmFTL is a soybean homolog of *FT* (Kong et al., 2010; Fan et al., 2014; Nan et al., 2014; Guo et al., 2015), a key controller for flowering initiation (Yoo et al., 2005; Corbesier et al., 2007; Tamaki et al., 2007; Andres and Coupland, 2012). Among them,

GmFTL3 and *GmFTL4* are strong candidates of *FT* genes in soybean (Kong et al., 2010; Fan et al., 2014). We previously reported that silencing *GmFTL* genes delays flowering in soybean (*Glycine max* cv. Tianlong1) (Guo et al., 2015). To investigate if *GmFTL* genes were involved in yield production, we selected four *GmFTL*-RNAi transgenic lines (RNAi line #1, RNAi line #3, RNAi line #4, and RNAi line #5) (Guo et al., 2015) for further study under controlled conditions. First, the expression levels of two major *FT* genes, *FTL3* and *GmFTL4*, in soybean leaves (Kong et al., 2010; Fan et al., 2014; Nan et al., 2014; Guo et al., 2015) were evaluated in a growth room. As **Supplementary Figure 1** shows, both genes displayed lower expression levels in transgenic lines. Among them, the abundance of *FTL3* and *GmFTL4* in either *GmFTL*-RNAi #4 or *GmFTL*-RNAi #5 was quite similar but was only one-tenth in wild-type plants, while *GmFTL*-RNAi#1 did not show a significant difference from wild-type plants. The levels of *GmFTL3/4* in *GmFTL*-RNAi#3 plants were between those of the wild-type and *GmFTL*-RNAi#4/5 plants. The results suggest that the effect of *GmFTL*-RNAi was much stronger on *GmFTL*-RNAi#4 and #5. Therefore, we focused on *GmFTL*-RNAi#4 as the main material, also combined with other lines in some cases, in the sequential studies under controlled conditions.

Larger shoots and roots of RNAi line #4 were easily observed (**Supplementary Figure 2**). The RNAi lines were taller with more nodes, and this phenotype was negatively correlated with the expression levels of *GmFTL3* and *GmFTL4* (**Supplementary Figure 3**). Therefore, these results suggest that *GmFTL* functions in promoting vegetative growth. However, RNAi line #3 had significant effect on branching and increased the number of branches, while other transgenic lines did not cause much difference in branching phenotypes compared with wild-type plants (**Supplementary Figure 3**).

There is no doubt, similar to other flowering plants, the flowering time and podding time are closely linked to the expression level of *FT* genes in soybean. Decreased *GmFTL3* and *GmFTL4* expression was attributed to late flowering (**Supplementary Figure 4**) and late podding (**Supplementary Figure 5**), suggesting a longer maturity phase for *GmFTL*-RNAi lines. Compared with wild-type plants, these *GmFTL*-RNAi lines displayed higher yield in both the growth room and greenhouse (**Figures 1A,B** and **Supplementary Figure 6**). RNAi line #1 also showed a slightly higher yield than wild-type plants, but this was not significant. The number of pods and seeds per plants likely contributed to the yield increase (**Supplementary Figure 7**) and such high yield did not occur at the cost of seed quality (**Supplementary Figure 8**). Obviously, higher intensity of light enhanced such yield-increasing effect, because the difference in yield between transgenic lines and wild-type plants was much greater in the greenhouse than in the growth room (**Figures 1A,B**, **Supplementary Figure 6**, and **Supplementary Table 1**). The length of light duration also had an obvious impact on yield, because the yield of WT and *GmFTL*-RNAi was higher in the greenhouse than in the growth room. In the growth room, the light/dark cycle is 8-h lighting from LEDs combined with 16-h dark; whereas in the greenhouse, the 24-h light/dark cycle has around 13-h lighting from sunlight with supplemental LEDs from 7:00 to 10:00 AM and 5:00 to 8:00 PM.

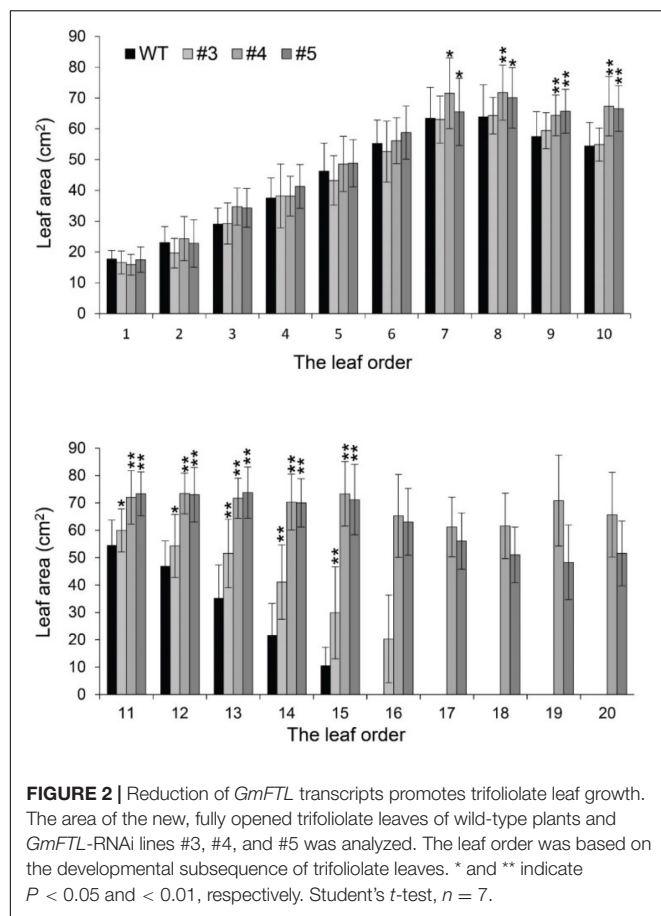


These results suggest that the duration of illumination affected the growth period, and that longer illumination prolonged the growth period, resulting in higher yield (**Supplementary Figure 6**). These results are consistent with typical characteristics of short-day plants, such as soybean (Bernier and Perilleux, 2005). It is also observed that even though the expression level of *GmFTL3* and *GmFTL4* and the flowering time were quite similar between transgenic lines #4 and #5, the difference in the yield increase was significant (**Supplementary Figures 1, 4, 6, 7**), suggesting that a minor change in the *GmFTL* transcript level would lead to a major change in yield. It should be noticed that a higher yield in *GmFTL*-RNAi lines #3, #4, and #5 occurred at the expense of a longer growth period (**Figure 1** and **Supplementary Figures 4–6**).

In *Arabidopsis* and rice, florigen interacts with a transcription factor FLOWERING LOCUS D (FD) (Taoka et al., 2011). We

also found that *GmFTL3* and *GmFTL4* proteins interacted with *GmFDL19* (**Supplementary Figure 9**), consistent with a previous report (Nan et al., 2014). No major perturbations in circadian expression of *GmFDL19* were observed in RNAi line #4 (**Figure 1D**). What is more, the circadian rhythm pattern of the *GmFTL3* and *GmFTL4* expression did not change (**Figure 1C**), which supports that the yield increase in *GmFTL*-RNAi lines results from the change in *GmFTL3* and *GmFTL4* genes at transcript abundance.

Florigen produces in leaves and then is transported to the apices to initiate flowers (Corbesier et al., 2007; Tamaki et al., 2007; Turck et al., 2008). *FD*, a functional partner of florigen, also has a potential function in leaves (Jang et al., 2017). Next, we confirmed that the effect of *GmFTL*-RNAi on yield was dominated by shoots or roots through a grafting approach experiment between *GmFTL*-RNAi line #4 and wild-type plants.



The results showed that the composite plants with *GmFTL*-RNAi shoots as a scion flowered later and had more seeds and larger roots than that with wild-type shoots as a scion (**Supplementary Figure 10**). However, the composite plants with wild-type plants as a scion had little effect on the related phenotypes. The results indicate that the shoots dominate the yield, flowering time, and root growth in *GmFTL*-RNAi line #4 plants.

Florigen Inhibits Leaf Growth

Beyond flowering regulation, florigen may be involved in many other development processes because it is expressed in many other tissues and organs besides leaf veins (Liu et al., 2014). A previous report proved that *FT* functions in leaf development (Teper-Bamnolker and Samach, 2005). We observed a visual phenotype that *GmFTL*-RNAi plants had greater number and sizes of leaves than wild-type plants (**Figure 2** and **Supplementary Figure 11**). A lower expression level of *FT* gene enhanced leaf growth, especially that of later initiated leaves (produced after the seventh trifoliolate leaves), which had a much larger size than early initiated ones, suggesting that the role of *GmFTL* in leaf growth and initiation is in a developmental stage-dependent mode. We further investigated the leaf structure in leaf sections. Transmission electron microscopy clearly showed that there was no significant difference in cell size of the early initiated leaves (the third trifoliolate leaves) between wild-type

and *GmFTL*-RNAi line #4 plants (**Figures 3A,B**). However, the cell size in the later initiated leaf (the seventh trifoliolate leaves) of *GmFTL*-RNAi line #4 was larger and longer (**Figures 3C,D**), suggesting that *GmFTL* was involved in leaf cell growth in a developmental stage-dependent mode in soybean. Previous studies have shown that florigen expresses increasingly according to developmental progress (Kardailsky et al., 1999; Krzymuski et al., 2015); and in that way, it is no surprising that the effect of *GmFTL*-RNAi is much obvious at late developmental stages. Taken together, the results suggest that *GmFTL* inhibits leaf growth and development in soybean.

Florigen Negatively Controls Photosynthesis

Next, we tried to elucidate the mechanism of *GmFTL* expression in soybean yield through transcriptome analysis of the third trifoliolate leaves of *GmFTL*-RNAi line #4 and WT at Zeitgeber1 (ZT1). Unexpectedly, RNA-seq data showed that silencing *GmFTL* caused expression change in only a small portion of coding genes in the soybean genome (0.378%, 212 out of 56,044, *G. max* Wm82.a2.v1) (**Supplementary Table 2**, differential expression genes, fold changes ≥ 2 and false discovery ≤ 0.05). Among them, chloroplast-related genes were highlighted (**Figures 4A,B**), suggesting that the effect of *GmFTL*-RNAi on leaves is limited and specific. Then, we selected genes related to chloroplast functions to confirm RNA-seq data by RT-qPCR. These genes included putative H^+ -ATP subunits (Glyma.11G110100, Glyma.14G151400, Glyma.06G067400, and Glyma.17G130100), putative NADH dehydrogenase subunits (Glyma.09G129000, Glyma.09G271600, and Glyma.10G128100), cytochrome b6f subunits (Glyma.15G114600 and Glyma.20G158300), and photosystem related genes (Glyma.06G224500 and Glyma.08G281300). The RT-qPCR results were in agreement with the RNA-seq data (**Figure 4C**). These data indicate that *GmFTL* may have a specific effect on chloroplast development and functions in soybean plants.

The structure of chloroplasts reflects the function of photosynthesis, and more thylakoid membranes and rich grana contribute to higher efficiency of photosynthesis and the formation of photosynthetic products (Jensen and Leister, 2014; Pribil et al., 2014; Kirchhoff, 2018). Therefore, we checked the characteristics of the chloroplast structure by transmission electron microscopy. The results showed that the chloroplasts of the *GmFTL*-RNAi#4 leaves exhibited much more complicated structures with more and wider thylakoid membranes and richer grana than those in wild-type plants, regardless of the early or later initiated leaves (**Figure 5**).

All the properties above may confer higher photosynthetic efficiency of *GmFTL*-RNAi line leaves. Then, we analyzed the physiological and biochemical characteristics of the leaves. The biochemical assay indicated that the *GmFTL*-RNAi line #4 leaves were enriched in photosynthetic pigments, had higher maximum quantum efficiency and photosynthetic rates (**Figure 6A**), and accumulated more photosynthetic assimilates such as starch, maltose, sucrose, glucose, and fructose (**Figure 6B**). Thus,

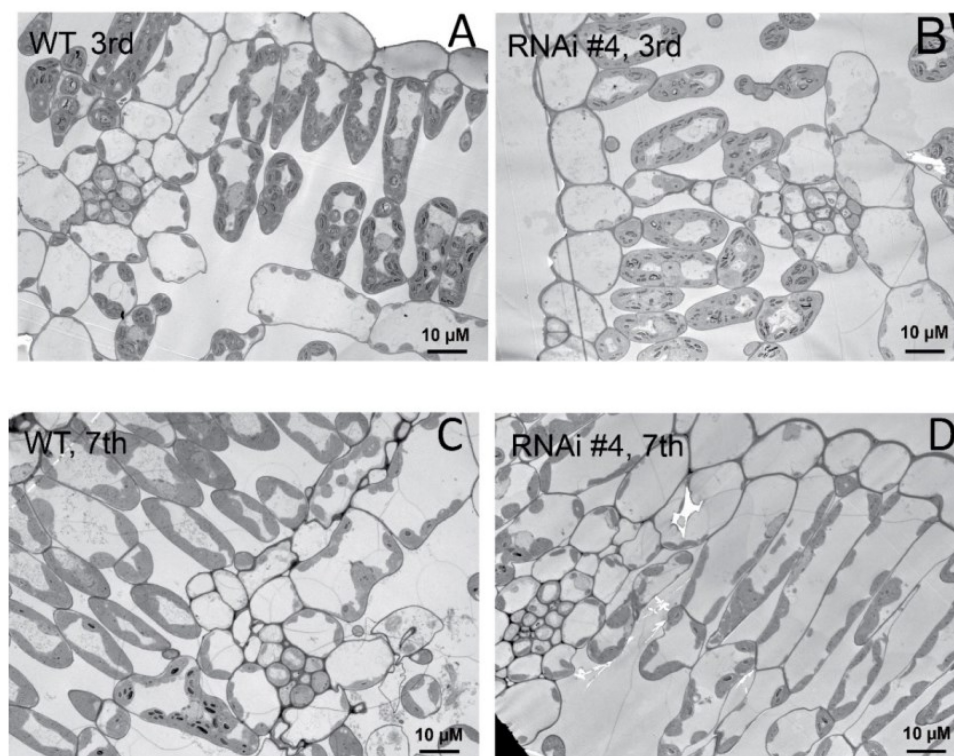


FIGURE 3 | *GmFTL*-RNAi enhances leaf cell growth. Transmission electron micrographs of the new, fully opened third (A,B) and seventh (C,D) trifoliolate leaves of wild-type (A,C) or the *GmFTL*-RNAi line #4 plants (B,D) grown in greenhouses. The seventh trifoliolate leaves show larger and longer cells in *GmFTL*-RNAi lines compared with wild-type plants. Scale bar, 10 μ m.

the transgenic *GmFTL*-RNAi plants had higher photosynthetic efficiency than the wild-type plants, and *GmFTL* negatively regulated photosynthesis.

A Slightly Decrease in Florigen Enhances Soybean Yield in the Field

All of the presented data above were from samples under controlled conditions. Then, we determined what would happen when these transgenic lines grew under natural field conditions. First, we planted *GmFTL*-RNAi lines #1, #3, and #4 in the field. Only *GmFTL*-RNAi line #1 matured naturally; and *GmFTL*-RNAi lines #3 and #4 did not mature before winter because they flowered too late. Therefore, we focused on *GmFTL*-RNAi line #1 for field experiments in two different environments (Beijing and Hanchuan) across years (from 2016 to 2018). In the field, there was not much difference between *GmFTL*-RNAi line #1 and wild-type plants during the vegetative stage (Supplementary Figure 12). However, at the fully mature stage, *GmFTL*-RNAi line #1 had more pods than wild-type plants (Supplementary Figure 13). Except for some cases of failure due to diseases, the yield increase in *GmFTL*-RNAi line #1 ranged from 7.2 to 24.2% (Figure 7). We also found that the yield increase in the original habitat (Hanchuan, N30°22', E113°22', where the parent line of *GmFTL*-RNAi line #1 originates from) was higher than that in the other environment (Beijing, N39°58', E116°20'). We postulated

that the altitude of growth regions restricts *GmFTL* functions, because soybean is an obligate short day and photoperiod-sensitive plant (Borthwick and Parker, 1938; Nanda and Hamner, 1959; Zhang et al., 2001). So, *GmFTL*-RNAi line #1 had at least an 11% yield increase compared with its parent in the original parent habitat (Figure 7B, 2018-Hanchuan). The result indicates that *GmFTL*-RNAi line #1 may be a high-yield elite candidate in the field, and that RNAi of florigen is a potential strategy to improve soybean yield.

DISCUSSION

Vegetative growth is a double-edged sword for reproductive growth: it is the foundation of reproductive growth, but an extended period of vegetative growth inhibits reproductive growth. Plants have evolved multiple strategies to balance these two essential processes, so that they can flower at an appropriate time and set enough healthy seeds to survive and prosper. Florigen, as a central activator of flowering time, may play a key role in balancing vegetative growth and reproductive growth. As a result of that, florigen is difficult to be applied in practical production, because its knockout leads to late flowering and failure of normal maturation in the field, whereas its overexpression results in early flowering and lower yield. The florigen dosage has a dominant role in regulating crop yield

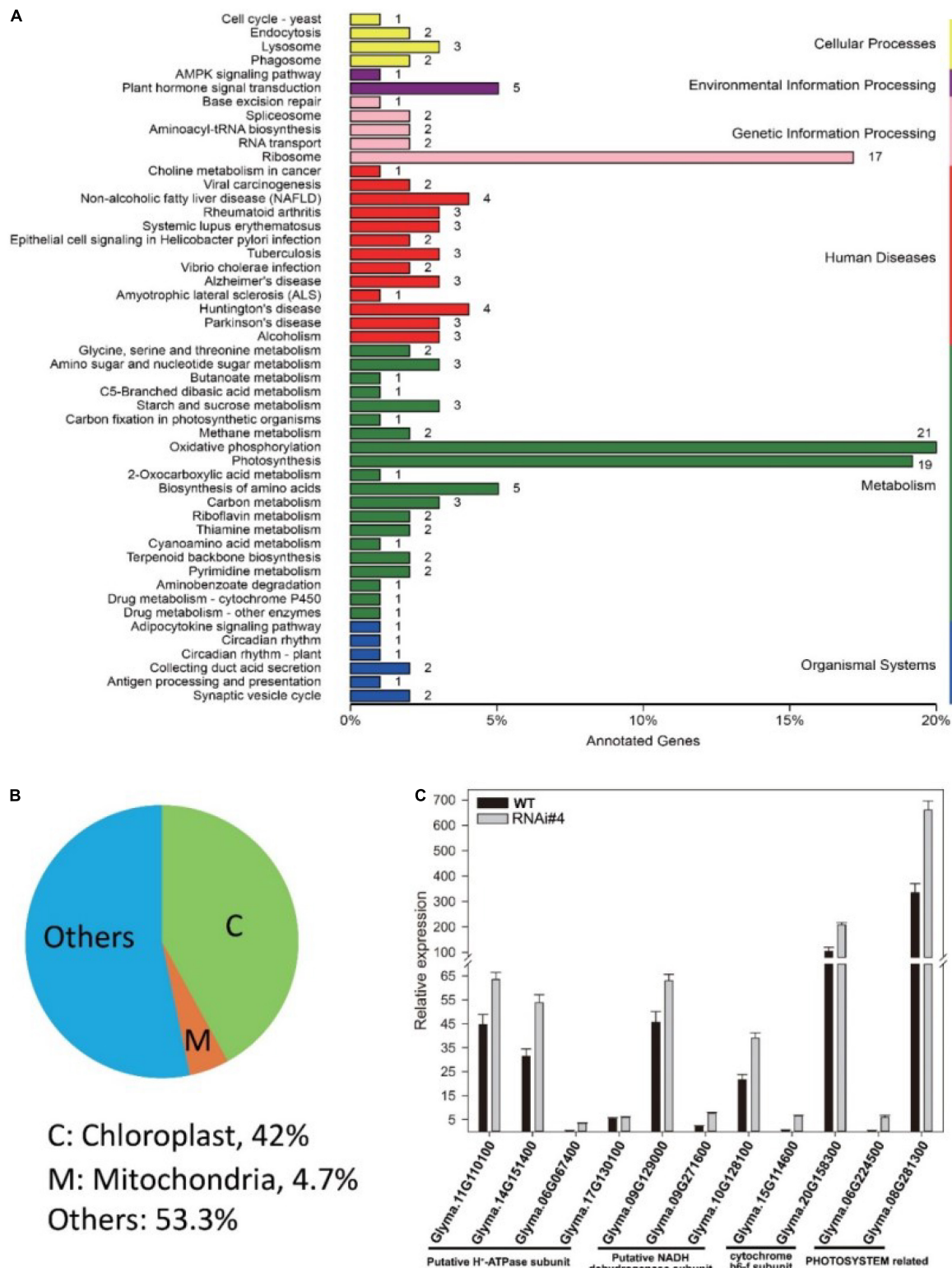


FIGURE 4 | *GmFTL*-RNAi significantly impacts the expression of genes related to photosynthesis. Transcriptome analysis of the third trifoliolate leaves of *GmFTL*-RNAi line #4 and WT at ZT1. **(A)** Differential expression genes (DEG, fold changes ≥ 2 and false discovery ≤ 0.05) were enriched in the KEGG pathways “Photosynthesis,” “Oxidative phosphorylation,” and “Ribosome,” which are mainly related to energy metabolism. **(B)** Forty-two percent of DEGs coded proteins targeting the chloroplast. Refer to **Supplementary Table 1** for the list of differential genes. **(C)** RT-qPCR verified the expression of 11 genes with more than two-fold changes detected by transcriptome analysis that are related to photosynthesis. The error bar indicates the standard deviation of three replicates.

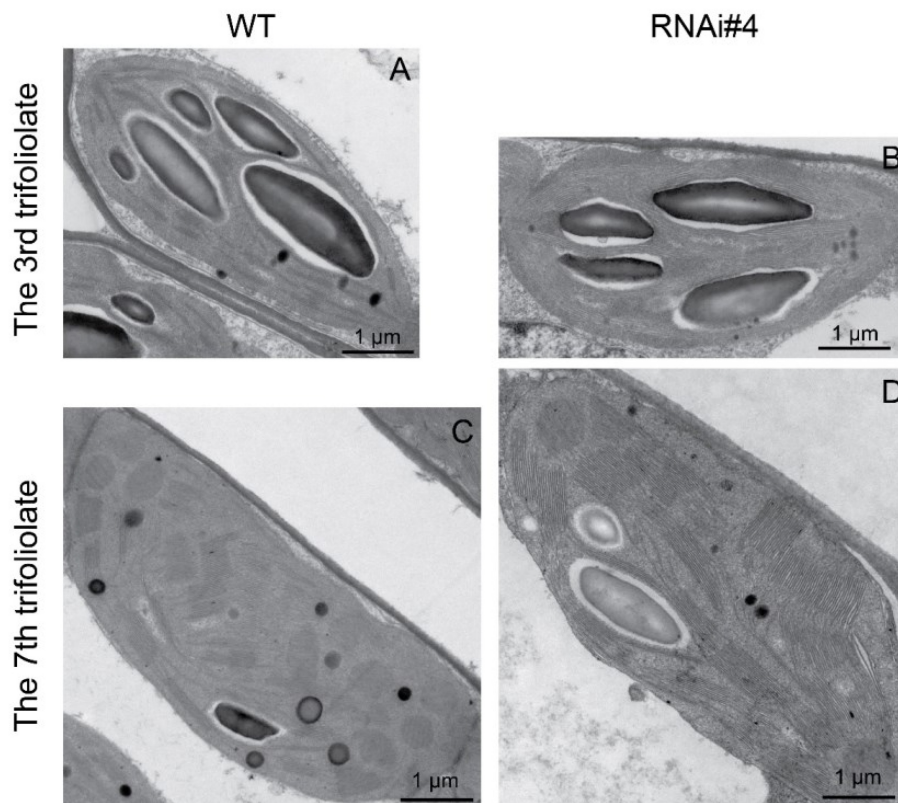


FIGURE 5 | *GmFTL*-RNAi enhances chloroplast development. The representative transmission electron micrographs of chloroplasts in the third (A,B) and seventh trifoliolate leaves (C,D) from wild-type plants (A,C) or *GmFTL* RNAi line #4 (B,D). Scale bar, 1 μ m.

(Krieger et al., 2010; Huang et al., 2016). However, how to utilize florigen to increase crop yield in the field remains unknown. In addition, there is no report showing that florigen is related to photosynthesis.

Florigen Inhibits Photosynthesis and Yield Production

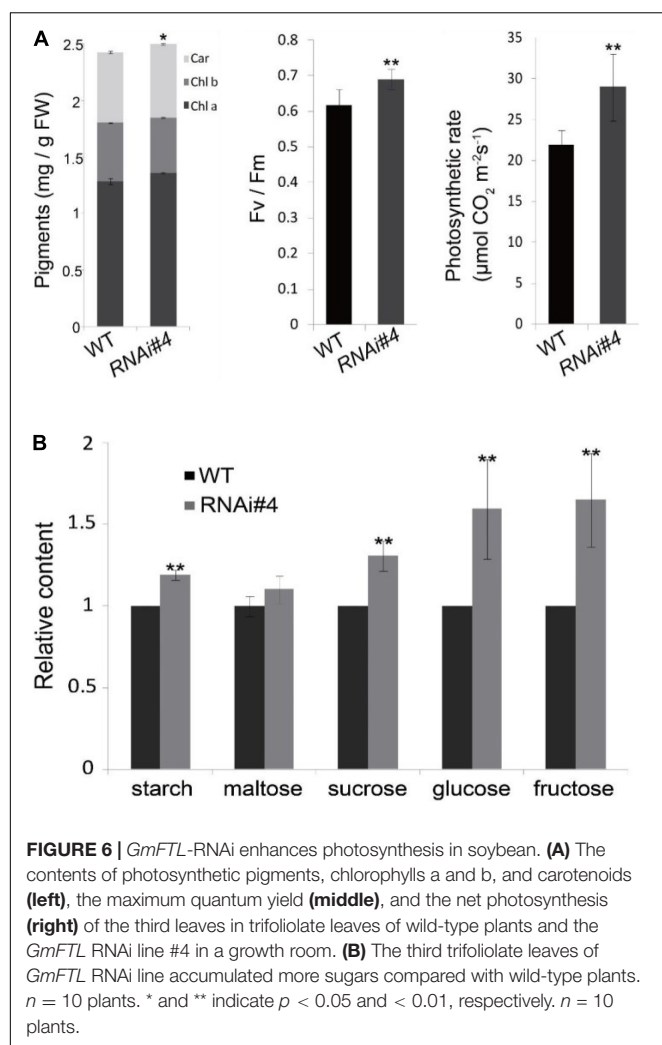
We demonstrate that the knockdown of florigen significantly enhances leaf growth and chloroplast development. *GmFTL*-RNAi lines have three typical advantages conferring high photosynthesis compared with wild-type plants: (1) more and larger leaves (Figure 2) with longer leaf cells (Figure 3); (2) chloroplasts with much more complicated membranes and grana (Figure 5); and (3) high amount of photosynthetic pigments (Figure 6). Such functions of *GmFTL*-RNAi are likely specific, because only few genes (212 out of 56,044 genes in the soybean genome, Supplementary Table 2) showed significant changes in expression, and 42% of these genes code proteins targeting chloroplasts (Figure 4). Therefore, it is no surprising that *GmFTL*-RNAi lines have more efficient photosynthesis, and they accumulate more photosynthetic products (Figure 6) and, finally, produce more seeds independent of environmental conditions (Figures 1, 7). However, growth conditions impact the function of *GmFTL*-RNAi, and higher intensity and longer duration of illumination enhance the effect (Figures 1, 7 and Supplementary

Table 1). The latitude of growth region is also linked to the function of *GmFTL*-RNAi; that is, the original habitat benefits the effect of *GmFTL*-RNAi (Figure 7), which can be tracked to native the habitat of a variety because soybean is an obligate short-day plant.

The *FT* gene is highly conserved in sequences and functions across the plant kingdom and plays multiple roles in many important processes beyond flowering time control (Pin et al., 2010; Navarro et al., 2011; Lee et al., 2013; Chen and Penfield, 2018). However, there is no study showing *FT* is related to photosynthesis, even though one study claims that *FT* has non-negligible functions in leaf growth in *Arabidopsis* (Teper-Bamnolker and Samach, 2005). This study bridges the gap between florigen and both photosynthesis and vegetative growth, which provides a cue to elucidate the mechanism of balancing vegetative growth and reproductive growth. It is interesting to identify the direct targets of *GmFTL* in leaves to establish a network coordinating vegetative growth and reproductive growth mediated by *GmFTL*.

Fine-Tuning Florigen Expression Is a Promising Strategy for High Yield

Crop yield attracts researchers because of the increasing global population. Numerous approaches have been employed to improve crop yield, such as manipulating the expression of



homologous (Preuss et al., 2012; Do et al., 2016; Ge et al., 2016) or heterologous genes in soybean (Li et al., 2013; Waltz, 2014; Kohler et al., 2016). Some strategies show high yield only under stress conditions (Li et al., 2013; Waltz, 2014; Do et al., 2016), whereas field trials have not been conducted (Waltz, 2014; Do et al., 2016; Ge et al., 2016). The approach of the authors shows soybean grain yield produced in 3 years in two field trials. Evaluation of the flowering time and yield of *GmFTL*-RNAi #1, compared with wild-type plants in the field or *GmFTL*-RNAi #4 compared with *GmFTL*-RNAi #5 under control conditions (Supplementary Figures 1, 6), revealed the importance of *GmFTL* level; that is, the yield difference would be significant even though flowering time is quite similar. Therefore, a slight reduction in the *FT* expression increases soybean yield in the field (Figure 7) but not at the expense of late flowering or seed quality (Supplementary Figures 1, 8), important traits for agricultural application (Blumel et al., 2015). The yield increase in the original habitat (Hanchuan) was 11–24% (Figure 7B). It is possible to screen more transgenic *GmFTL*-RNAi lines with different florigen levels to obtain a much higher yield in the field.

Different environmental conditions affect the yield of *GmFTL*-RNAi lines (Figures 1, 7), indicating that environmental cues, such as light characteristics, participate in the network of *GmFTL*-RNAi regulation of yield (or photosynthesis). Therefore, it is interesting to develop a new strategy for modifying the light signaling pathway to increase soybean yield.

In this study, we employed one key gene, florigen, and common biotechnology, RNAi. Florigen is a highly conserved gene in the plant kingdom that enhances plant flowering (Wickland and Hanzawa, 2015). RNAi is a classical approach to reducing gene expression. To confirm such results, we also analyzed the *FT*-RNAi effect on leaf development in *Arabidopsis thaliana*. As expected, the number of rosette leaves was inversely proportional to the abundance of *FT* mRNA (Supplementary Figure 14), and *FT*-RNAi increased the biomass and the size of leaves of *Arabidopsis*. Therefore, we predict that the strategy we have shown here can be widely applied to different crops to increase their yields in the field. Beyond florigen, many genes are involved in the regulation of flowering time, and similar challenges will be met when these flowering genes are applied in agricultural production. The strategy provides a smart example for the application of such genes.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

GmFTL-RNAi plants were previously generated in the soybean [*G. max* (L.) Mer.] cultivar Tianlong1 in the laboratory of the authors (Guo et al., 2015). All the RNAi lines used here are independent, homozygous, transgenic ones. The wild-type soybean control for all of the experiments is cultivar Tianlong1, which originates in Hanchuan, China. All of the plants were grown under controlled temperature and photoperiod growth rooms, greenhouses, and/or the field. The light conditions in the plant growth rooms are short day conditions (8 h light/16 h dark) from a LED light source (GreenPower LED top lighting, Philips Horticulture LED).¹ Natural light and temperature conditions were used in the greenhouse, with supplemental LED light from 7:00 to 10:00 AM and 5:00 to 8:00 PM. The light spectrum and intensity in the growth room and greenhouse are listed in Supplementary Table 1.

Soybean field (plot) experiments were carried out in Hanchuan (N30°22', E113°22') and Beijing (N39°58', E116°20') from 2016 to 2018. Spring sown soybeans were planted on 23rd April 2016 for Hanchuan, and May 15 or June 1, 2017 for Beijing. The plot area was 300 cm × 225 cm with 45-cm row-spacing and 20-cm plant-spacing in Hanchuan or 600 cm × 300 cm with 60-cm row-spacing and 30 cm plant-spacing in Beijing. Agronomic characters were determined at maturity. Individual plants from each plot were subjected to statistical analysis.

RNA Extraction and Expression Analysis

Total RNA of leaves was extracted using an EasyPure® RNA Kit (ER101-01, TransGen Biotech, Beijing, China). The quantity

¹<http://www.lighting.philips.com.cn/products/horticulture>

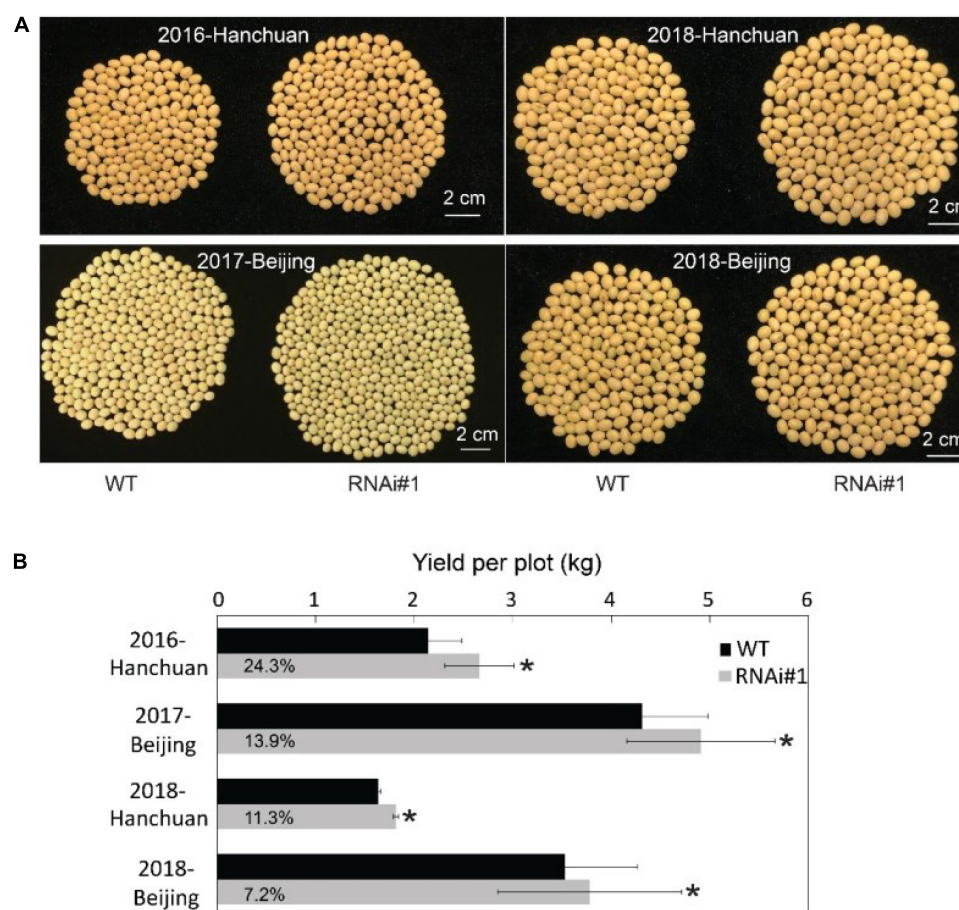


FIGURE 7 | Slightly reducing *GmFTL* expression increases soybean yield in the field. **(A)** *GmFTL*-RNAi line #1 was subjected to a field yield investigation in Hanchuan and Beijing from 2016 to 2018. The photos show the total number of seeds from a single representative plant. **(B)** Statistical assay of yield per plot (kg) in the field investigation. The digits in the gray bars indicate the percentage of yield increase in *GmFTL*-RNAi#1 over wild-type plants. *Indicates significant differences at $P < 0.05$ based on Student's *t*-test. $n = 5$.

was measured with Nanodrop 2000C (Thermo Fisher Scientific, Waltham, MA, United States). For SYBR detection of RT-qPCR products, 500–1,000 ng of the total RNA were used for reverse transcription (KR106-02, TIANGEN, Sichuan, China). SYBR Premix Ex-Taq (Perfect Real Time; TaKaRa, Tokyo, Japan) was used for the RT-qPCR assays. For TaqManTM analysis, about 1 μ g of total RNA was used for reverse transcription (KR106-02, TIANGEN, Sichuan, China), and TaqMan Gene Expression Master Mix (No.4369016, Thermo Fisher Scientific, Waltham, MA, United States) was used for the assays. RT-qPCR was conducted using StepOne Plus (ABI, United States). The reference gene *GmACT11* (Glyma.18G290800) or *GmUKN2* (Glyma06g04180) (Hu et al., 2009) was used as internal control. Sequences of the primers are listed in **Supplementary Table 3**. The $2^{-\Delta CT}$ method was used to calculate the relative expression levels based on three technical replicates.

Bimolecular Fluorescence Complementation

The pEarlyGate202-*GmFDL19-cYFP* and pEarlyGate202-*GmFTL3-nYFP* or pEarlyGate201-*GmFTL4-nYFP* binary vectors

were transiently expressed using *Agrobacterium tumefaciens*. Both recombinant *Agrobacterium* cells were co-infiltrated into *Nicotiana benthamiana* leaves. Empty vectors were used as negative controls, and AtAHL22-RFP was used as a nuclear marker (Xiao et al., 2009). *N. benthamiana* was grown under long-day (16-h light:8-h dark) conditions at 22°C for at least 48-h post infiltration. Leaves were observed under a confocal microscope (Zeiss LSM700, Jena, Germany).

Chloroplast Analysis and Chlorophyll Measurements

The new, fully opened third and seventh trifoliolate leaves of wild-type plants Tianlong1 and *GmFTL*-RNAi line #4 soybean plants in the greenhouse were harvested for measurements. Perpendicular transverse sections of the middle leaflets of trifoliolate leaves were prepared by the Transmission-Electron-Microscope and Mass-Spectrometry Platform of Institute of Agricultural Products Processing, Chinese Academy of Agricultural Sciences (Beijing, China). The photographs were obtained using a transmission electron microscope (H-7500, Hitachi, Tokyo, Japan).

For chlorophyll content analysis, we punched 20 fresh sections ($d = 6$ mm) from five individual leaves of each sample using a hole puncher. The samples were immersed in 25 mL of 80% acetone and stored at room temperature for 5 days in the dark. Then, the absorbance of 1 ml of the supernatants was measured at 663 and 645 nm. The concentration of chlorophylls a and b (Chl a and Chl b) and carotenoids was calculated using the following formulas:

$$\text{Chl a} = (11.24A_{663} - 2.04A_{645}) \times V/W$$

$$\text{Chl b} = (20.13A_{645} - 4.19A_{663}) \times V/W$$

Carotenoids

$$= ((1000A_{470} - 1.9\text{Chl a} - 63.14\text{Chl b})/214) \times V/W$$

(V is the volume of acetone, and W is the fresh weight of the sample)

Photosynthetic Rate and Chlorophyll Fluorescence Analyses

The plants were grown under SD conditions (growth room), and the third trifoliolate leaves ($n = 10$) of wild-type plants Tianlong1 and *GmFTL*-RNAi line #4 were selected to measure the photosynthetic rates at days 28–35 after sowing following the instructions of the manufacturer (LI-6400 V4.0.1, LI-COR, Lincoln, NE, United States). Fv/Fm was measured using an IMAGING-PAM M series Chlorophyll Fluorometer with the MAXI version (Heinz Walz, Effeltrich, Germany). The plants were placed in the dark for 30 min before measurements were taken.

Measurement of the Sugar Content in Leaves

Soybean Tianlong1 and *GmFTL*-RNAi line 4 seeds were sown in a growth room. The third new, fully opened trifoliolate leaves were harvested at ZT1 for the measurement of sugar content ($n = 5$). For the extraction of soluble sugars and starch, about 100 mg of leaf samples was homogenized in 5 ml of 80% (v/v) ethanol in a 1.5-ml tube and incubated at 70°C for 90 min. Following centrifugation at $16,000 \times g$ for 5 min, the supernatant was transferred to a new 1.5-ml tube. The pellet was rinsed twice with 2 ml of 80% ethanol and removed. Any remaining solvent was evaporated at room temperature using a vacuum. The residue was resuspended in 3 ml of distilled, sterile water, and this represented the soluble carbohydrate fraction. The remaining pellet containing insoluble carbohydrates, such as starch, was homogenized in 2 ml of 2 N KOH, and the suspension was incubated at 95°C for 1 h to dissolve the starch. Following the addition of 0.35 ml of 1 N acetic acid and centrifugation for 5 min at $16,000 \times g$, the supernatant was used for starch quantification. Detailed procedures were followed according to the instructions of the manufacturer; starch (No. 1013910603), maltose/sucrose/D-glucose (No. 11113950035) and D-glucose/D-fructose kits (No. 10139106035) (R-Biopharm, Pfungstadt, Germany).

Measurement of Major Agronomic Traits

The flowering time was determined by the emergence of the first flower on the main stem of soybean plants. The podding time was determined as the time when the first pod on the main stem of the soybean plants was 2 cm in length. Measurements of plant height, branch number, node number, pod number and seed number per plant were performed at full plant maturity. For experiments in both the greenhouse and growth rooms, more than 10 individual plants of each line were sampled for analysis of all traits. For analysis of seed quality traits, 100 plump seeds of wild-type plants and transgenic lines were used. The content of proteins, oils, and moisture was determined using a near-infrared spectrometer (Brooke Technology Co., Ltd., Beijing, China) with the pre-stored soybean protein and oil model in the instrument. The operation was performed according to the instructions of the manufacturer.

Transcriptome Analysis

The new, fully opened third trifoliolate leaves from wild-type plants or *GmFTL*-RNAi line #4 were sampled ZT1. Total RNA was extracted using an EasyPure® RNA Kit (ER101-01, TransGen Biotech, Beijing, China). RNA sequencing with an Illumina HiSeq instrument and data analysis were performed by Biomarker Technologies (Beijing, China). Illumina sequencing reads were mapped to reference genome *G. max* Wm82.a2.v1.² Transcriptome data were deposited to Genome Sequence Archive (GSA)³ with an accession number of CRA004267).

Statistical Analysis

All experiments in this study were carried out with at least three replicates, all of which showed similar results. The figures showed only a representative result. Data in all bar graphs represent the mean \pm SD. All statistical analyses were performed using the SPSS software package. Asterisks indicate significant difference based on a Student's *t*-test (** $P < 0.01$, * $P < 0.05$).

Gene Accession Numbers

Sequence data for this article can be found in the Phytozome: *GmFTL3* (Glyma16g26660), *GmFTL4* (Glyma16g04830), *GmFDL19* (Glyma.19G122800), *GmACT11* (Glyma.18G290800), *GmUKN2* (Glyma06g04180), putative H⁺-ATP subunits (Glyma.11G110100, Glyma.14G151400, Glyma.06G067400, and Glyma.17G130100), putative NADH dehydrogenase subunits (Glyma.09G129000, Glyma.09G271600, and Glyma.10G128100), cytochrome b6f subunits (Glyma.15G114600 and Glyma.20G158300), photosystem related genes (Glyma.06G224500 and Glyma.08G281300), and *AtFT* (At1g65480).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: <https://ngdc.cnbc.ac.cn/gsa/browse/CRA004267>.

²<https://phytozome.jgi.doe.gov/pz/portal.htm>

³<https://ngdc.cnbc.ac.cn/gsub/>

AUTHOR CONTRIBUTIONS

Y-FF, X-MZ, and XF: conceptualization. KX, Y-FF, X-MZ, CZ, and HC: methodology. KX, X-MZ, HC, JZ, CZ, ZC, PH, and YM: investigation. KX, X-MZ, and Y-FF: formal analysis and validation. KX and Y-FF: visualization. Y-FF, KX, and XF: writing. XF, X-MZ, and Y-FF: funding acquisition. Y-FF and X-MZ: project administration. XZ, Y-FF, and YM: resources. Y-FF: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the National Transgenic Major Project of China (Grant No: 2016ZX08004-005), the National Key R&D Project (Grant No: 2016YFD0101900) from the Ministry of Science and Technology of China, the National Natural Science Foundation of China (Grant Nos: 31771714, 31371703 and 31570289), and the CAAS-Innovation Team Project and the Basal Research Fund of CAAS (Y2017CG25).

ACKNOWLEDGMENTS

We wish to thank all the friends and colleagues who helped us in any way but are not included in the author list.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.710754/full#supplementary-material>

Supplementary Figure 1 | *GmFTL*-RNAi reduces the mRNA abundance of *GmFTL3* and *GmFTL4* in different transgenic lines. WT and *GmFTL*-RNAi lines #1, #3, #4, and #5 grew in growth room, and the first trifoliolate leaves were harvested to investigate gene expression at ZT4 by RT-qPCR. *GmACT11* was used as a reference gene. Among these transgenic lines, line #1 shows slight change in *GmFTL3* and *GmFTL4* expressions. Error bars indicate the standard deviation of the mean of three replicates. An asterisk indicates significant difference compared with wild-type plant (**, $P < 0.01$. Student's *t*-test, $n \geq 5$ plants).

Supplementary Figure 2 | Reducing *GmFTL* expression enhances the growth of (A) shoots and (B) roots in soybean. WT and *GmFTL*-RNAi line #4 grown in greenhouse. Photos were taken at maturity of wild-type plants.

Supplementary Figure 3 | Reducing *GmFTL* expression enhances the growth of (A,B) stems but not (C) branching. WT and *GmFTL*-RNAi line #4 grown in greenhouse. Agronomic traits were analyzed at maturity. An asterisk indicates significant difference compared with wild-type plants (** $P < 0.01$. Student's *t*-test, $n \geq 20$ plants).

Supplementary Figure 4 | Reducing *GmFTL* expression delays soybean flowering. WT and *GmFTL*-RNAi lines #1, #3, #4, and #5 grown in greenhouse. An asterisk indicates significant difference compared with wild-type plants (** $P < 0.01$. Student's *t*-test, $n \geq 20$ plants).

Supplementary Figure 5 | Reducing *GmFTL* expression delays soybean podding time. WT and *GmFTL*-RNAi lines #1, #3, #4, and #5 grown in

greenhouse. An asterisk indicates significant difference compared with wild-type plants (** $P < 0.01$. Student's *t*-test, $n \geq 20$ plants).

Supplementary Figure 6 | Reducing *GmFTL* expression increases soybean yield. WT and *GmFTL*-RNAi lines #1, #3, #4, and #5 grown in growth room and greenhouse. This figure is supporting data for **Figure 1**. An asterisk indicates significant difference compared with wild-type plants (** $P < 0.01$. Student's *t*-test, $n \geq 10$ plants).

Supplementary Figure 7 | Reducing *GmFTL* expression increases the number of pods and seeds per plant. WT and *GmFTL*-RNAi lines #1, #3, #4, and #5 grown in greenhouse. An asterisk indicates significant difference compared with wild-type plants (** $P < 0.01$. Student's *t*-test, $n \geq 20$ plants).

Supplementary Figure 8 | Reducing *GmFTL* expression does not alter the quality of soybean seeds. *GmFTL*-RNAi lines #4 and #5 grown in greenhouse have similar contents of (A) proteins, (B) oils, and (C) water as wild-type seeds.

Supplementary Figure 9 | Both *GmFTL3* and *GmFTL4* proteins interact with *GmFDL19* in the nucleus. *Agrobacterium tumefaciens* harboring pEarlyGate201-*GmFTL3::nYFP*, or pEarlyGate201-*GmFTL4::nYFP*, pEarlyGate202-*GmFDL19::cYFP* vectors were co-infiltrated into *Nicotiana benthamiana* leaves, respectively. Then, plants were incubated for 48 h and subsequently observed by bimolecular fluorescence complementation (BiFC) under a confocal microscope. AHL22-RFP was used as a nucleus marker protein. *GmFTL3::nYFP* and cYFP, *GmFTL4::nYFP* and cYFP, and nYFP and *GmFDL19::cYFP* were used as controls.

Supplementary Figure 10 | Grafting assay. (A) Photos of different graft combinations and ungrafted control plants at maturity. For each graft combination, the genotypes of scion (above) and stock (below) were separated by a red line. (B) Flowering time and (C) seed number per plant for different graft combinations. Statistical significance at the 0.01 level was determined by Duncan's multiple range test, $n \geq 3$ plants.

Supplementary Figure 11 | Reduction of *GmFTL* transcripts promotes growth of trifoliolate leaves. This photo shows morphology of trifoliolate leaves of *GmFTL*-RNAi #3, #4, and #5 grown in greenhouse. The leaf order is according to the developmental subsequence of trifoliolate leaves. This figure is a supporting data for **Figure 4**.

Supplementary Figure 12 | *GmFTL*-RNAi line #1 does not show significant difference at the vegetative stage compared with wild-type plants. This photo shows plants of *GmFTL*-RNAi #1 grown in the field in Hanchuan, the original location of wild-type plants, in 2018. The photo indicates there is no difference between *GmFTL*-RNAi #1 (right) and wild-type (left) plants. Seeds were sown on April 20 and photos were taken on June 5.

Supplementary Figure 13 | *GmFTL*-RNAi line #1 shows significant difference at the reproductive stage from wild-type plants. This photo shows dry plants at maturity of *GmFTL*-RNAi #1 and wild-type plants grown in the field in Hanchuan, the original location of wild-type plants, in 2018. The photo displays that *GmFTL*-RNAi #1 plants (low row) are of bigger stature and have more pods than wild-type plants (up row).

Supplementary Figure 14 | The expression level of florigen is positively related to biomass and leaf size in *Arabidopsis*. *AtFT*-RNAi lines were produced by introducing an RNAi fragment of the *Arabidopsis* florigen *FT* gene (At1g65480) into ecotype Columbia-0, and homozygous transgenic lines were used in all experiments. *ft-10* is a null mutant of the florigen *AtFT* gene. *Arabidopsis* plants (wild-type, *AtFT*-RNAi lines, and *ft-10* mutants) grew under long day conditions (growth room). Photos of (A) rosettes and (B) leaves were taken at flowering. The *Arabidopsis* florigen *FT* gene expression was determined by RT-qPCR at day 14 after germination (C). The flowering time is negatively correlated to flowering time (C) under long day conditions.

Supplementary Table 1 | Light information in growth room and greenhouse.

Supplementary Table 2 | Transcriptome analysis of *GmFTL*-RNAi#4 leaves at ZT1.

Supplementary Table 3 | List of oligonucleotide and primer sequences used in this study.

REFERENCES

- Andres, F., and Coupland, G. (2012). The genetic basis of flowering responses to seasonal cues. *Nat. Rev. Genet.* 13, 627–639. doi: 10.1038/nrg3291
- Bernier, G., and Perilleux, C. (2005). A physiological overview of the genetics of flowering time control. *Plant Biotechnol. J.* 3, 3–16. doi: 10.1111/j.1467-7652.2004.00114.x
- Blumel, M., Dally, N., and Jung, C. (2015). Flowering time regulation in crops—what did we learn from Arabidopsis? *Curr. Opin. Biotechnol.* 32, 121–129. doi: 10.1016/j.copbio.2014.11.023
- Borthwick, H. A., and Parker, M. W. (1938). Photoperiodic perception in Biloxi soybeans. *Bot. Gaz.* 100, 374–387. doi: 10.1086/334792
- Chen, M., and Penfield, S. (2018). Feedback regulation of COOLAIR expression controls seed dormancy and flowering time. *Science* 360, 1014–1017. doi: 10.1126/science.aar7361
- Cho, L. H., Yoon, J., and An, G. (2017). The control of flowering time by environmental factors. *Plant J.* 90, 708–719. doi: 10.1111/tpj.13461
- Corbesier, L., Vincent, C., Jang, S., Fornara, F., Fan, Q., Searle, I., et al. (2007). FT protein movement contributes to long-distance signaling in floral induction of Arabidopsis. *Science* 316, 1030–1033. doi: 10.1126/science.1141752
- Do, T. D., Chen, H., Hien, V. T., Hamwieh, A., Yamada, T., Sato, T., et al. (2016). Ncl synchronously regulates Na(+), K(+), and Cl(-) in Soybean and greatly increases the grain yield in saline field conditions. *Sci. Rep.* 6: 19147.
- Fan, C., Hu, R., Zhang, X., Wang, X., Zhang, W., Zhang, Q., et al. (2014). Conserved CO-FT regulons contribute to the photoperiod flowering control in soybean. *BMC Plant Biol.* 14:9. doi: 10.1186/1471-2229-14-9
- Ge, L., Yu, J., Wang, H., Luth, D., Bai, G., Wang, K., et al. (2016). Increasing seed size and quality by manipulating BIG SEEDS1 in legume species. *Proc. Natl. Acad. Sci. U. S. A.* 113, 12414–12419. doi: 10.1073/pnas.1611763113
- Guo, G., Xu, K., Zhang, X., Zhu, J., Lu, M., Chen, F., et al. (2015). Extensive analysis of GmFTL and GmCOL expression in Northern Soybean cultivars in field conditions. *PLoS One* 10:e0136601. doi: 10.1371/journal.pone.0136601
- Hu, R., Fan, C., Li, H., Zhang, Q., and Fu, Y. F. (2009). Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR. *BMC Mol. Biol.* 10:93. doi: 10.1186/1471-2199-10-93
- Huang, X., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., et al. (2016). Genomic architecture of heterosis for yield traits in rice. *Nature* 537, 629–633. doi: 10.1038/nature19760
- Jang, S., Li, H. Y., and Kuo, M. L. (2017). Ectopic expression of Arabidopsis FD and FD PARALOGUE in rice results in dwarfism with size reduction of spikelets. *Sci. Rep.* 7:44477.
- Jensen, P. E., and Leister, D. (2014). Chloroplast evolution, structure and functions. *F1000Prime Rep.* 6:40.
- Kardailsky, I., Shukla, V. K., Ahn, J. H., Dagenais, N., Christensen, S. K., Nguyen, J. T., et al. (1999). Activation tagging of the floral inducer FT. *Science* 286, 1962–1965. doi: 10.1126/science.286.5446.1962
- Kirchhoff, H. (2018). Structure-function relationships in photosynthetic membranes: challenges and emerging fields. *Plant Sci.* 266, 76–82. doi: 10.1016/j.plantsci.2017.09.021
- Kohler, I. H., Ruiz-Vera, U. M., VanLoocke, A., Thomey, M. L., Clemente, T., Long, S. P., et al. (2016). Expression of cyanobacterial FBP/SBPase in soybean prevents yield depression under future climate conditions. *J. Exp. Bot.* 68, 715–726.
- Kong, F., Liu, B., Xia, Z., Sato, S., Kim, B. M., Watanabe, S., et al. (2010). Two coordinately regulated homologs of FLOWERING LOCUS T are involved in the control of photoperiodic flowering in soybean. *Plant Physiol.* 154, 1220–1231. doi: 10.1104/pp.110.160796
- Krieger, U., Lippman, Z. B., and Zamir, D. (2010). The flowering gene single Flower Truss drives heterosis for yield in tomato. *Nat. Genet.* 42, 459–463. doi: 10.1038/ng.550
- Krzyszowski, M., Andres, F., Cagnola, J. I., Jang, S., Yanovsky, M. J., Coupland, G., et al. (2015). The dynamics of Flowering Locus T expression encodes long-day information. *Plant J.* 83, 952–961. doi: 10.1111/tpj.12938
- Lee, R., Baldwin, S., Kenel, F., McCallum, J., and Macknight, R. (2013). FLOWERING LOCUS T genes control onion bulb formation and flowering. *Nat. Commun.* 4:2884.
- Li, Y., Zhang, J., Hao, L., Hua, J., Duan, L., Zhang, M., et al. (2013). Expression of an Arabidopsis molybdenum cofactor sulphurase gene in soybean enhances drought tolerance and increases yield under field conditions. *Plant Biotechnol. J.* 11, 747–758. doi: 10.1111/pbi.12066
- Liu, L., Farrona, S., Klemme, S., and Turck, F. K. (2014). Post-fertilization expression of FLOWERING LOCUS T suppresses reproductive reversion. *Front. Plant Sci.* 5:164.
- Long, S. P., Marshall-Colon, A., and Zhu, X. G. (2015). Meeting the global food demand of the future by engineering crop photosynthesis and yield potential. *Cell* 161, 56–66. doi: 10.1016/j.cell.2015.03.019
- Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330. doi: 10.1111/j.1365-3040.2005.01493.x
- Nan, H., Cao, D., Zhang, D., Li, Y., Lu, S., Tang, L., et al. (2014). GmFT2a and GmFT5a redundantly and differentially regulate flowering through interaction with and upregulation of the bZIP transcription factor GmFDL19 in soybean. *PLoS One* 9:e97669. doi: 10.1371/journal.pone.0097669
- Nanda, K. K., and Hamner, K. C. (1959). Photoperiodic cycles of different lengths in relation to flowering in Biloxi soybean. (Glycine Max L. Merr.). *Planta* 53, 45–52. doi: 10.1007/bf02109487
- Navarro, C., Abelenda, J. A., Cruz-Oro, E., Cuellar, C. A., Tamaki, S., Silva, J., et al. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature* 478, 119–122. doi: 10.1038/nature10431
- Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8529–8536. doi: 10.1073/pnas.1424031112
- Parry, M. A. J., Madgwick, P. J., Carvalho, J. F. C., and Andralojc, P. J. (2007). Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J. Agric. Sci.* 145, 31–43. doi: 10.1017/s0021859606006666
- Pin, P. A., Benlloch, R., Bonnet, D., Wremeth-Weich, E., Kraft, T., Gielen, J. J., et al. (2010). An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. *Science* 330, 1397–1400. doi: 10.1126/science.1197004
- Preuss, S. B., Meister, R., Xu, Q., Urwin, C. P., Tripodi, F. A., Screen, S. E., et al. (2012). Expression of the *Arabidopsis thaliana* BBX32 gene in soybean increases grain yield. *PLoS One* 7:e30717. doi: 10.1371/journal.pone.0030717
- Pribil, M., Labs, M., and Leister, D. (2014). Structure and dynamics of thylakoids in land plants. *J. Exp. Bot.* 65, 1955–1972. doi: 10.1093/jxb/eru090
- South, P. F., Cavanagh, A. P., Liu, H. W., and Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* 363:eaat9077. doi: 10.1126/science.aat9077
- Tamaki, S., Matsuo, S., Wong, H. L., Yokoi, S., and Shimamoto, K. (2007). Hd3a protein is a mobile flowering signal in rice. *Science* 316, 1033–1036. doi: 10.1126/science.1141753
- Taoka, K., Ohki, I., Tsuji, H., Furuita, K., Hayashi, K., Yanase, T., et al. (2011). 14-3-3 proteins act as intracellular receptors for rice Hd3a florigen. *Nature* 476, 332–U397.
- Teper-Bamnolker, P., and Samach, A. (2005). The flowering integrator FT regulates SEPALLATA3 and FRUITFULL accumulation in Arabidopsis leaves. *Plant Cell* 17, 2661–2675. doi: 10.1105/tpc.105.035766
- Tilman, D., Balzer, C., Hill, J., and Belfort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20260–20264.
- Turck, F., Fornara, F., and Coupland, G. (2008). Regulation and identity of florigen: FLOWERING LOCUS T moves center stage. *Annu. Rev. Plant Biol.* 59, 573–594. doi: 10.1146/annurev.arplant.59.032607.092755
- Waltz, E. (2014). Beating the heat. *Nat. Biotechnol.* 32, 610–613. doi: 10.1038/nbt.2948
- Wickland, D. P., and Hanzawa, Y. (2015). The FLOWERING LOCUS T/TERMINAL FLOWER 1 gene family: functional evolution and molecular mechanisms. *Mol. Plant* 8, 983–997. doi: 10.1016/j.molp.2015.01.007
- Xiao, C., Chen, F., Yu, X., Lin, C., and Fu, Y. F. (2009). Over-expression of an AT-hook gene, AHL22, delays flowering and inhibits the elongation of the

- hypocotyl in *Arabidopsis thaliana*. *Plant Mol. Biol.* 71, 39–50. doi: 10.1007/s11103-009-9507-9
- Yoo, S. K., Chung, K. S., Kim, J., Lee, J. H., Hong, S. M., Yoo, S. J., et al. (2005). CONSTANS activates SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1 through FLOWERING LOCUS T to promote flowering in *Arabidopsis*. *Plant Physiol.* 139, 770–778. doi: 10.1104/pp.105.066928
- Zhang, L., Wang, R., and Hesketh, J. D. (2001). Effects of photoperiod on growth and development of soybean floral bud in different maturity. *Agron. J.* 93, 944–948. doi: 10.2134/agronj2001.934944x

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 © 2021 Xu, Zhang, Chen, Zhang, Zhu, Cheng, Huang, Zhou, Miao, Feng and Fu. 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 Transcriptomic and Bioinformatics Analyses Reveal the Molecular Mechanisms for the Differences in Seed Oil and Starch Content Between *Glycine max* and *Cicer arietinum*

Kun Cheng, Yi-Fan Pan, Lü-Meng Liu, Han-Qing Zhang and Yuan-Ming Zhang*

Crop Information Center, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Chaofu Lu,
Montana State University,
United States
Dan Zhang,
Henan Agricultural University, China

*Correspondence:

Yuan-Ming Zhang
soyzzhang@mail.hzau.edu.cn

Specialty section:

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

Received: 19 July 2021

Accepted: 23 September 2021

Published: 26 October 2021

Citation:

Cheng K, Pan Y-F, Liu L-M,
Zhang H-Q and Zhang Y-M (2021)
Integrated Transcriptomic and
Bioinformatics Analyses Reveal the
Molecular Mechanisms for the
Differences in Seed Oil and Starch
Content Between *Glycine max* and
Cicer arietinum.
Front. Plant Sci. 12:743680.
doi: 10.3389/fpls.2021.743680

The seed oil and starch content of soybean are significantly different from that of chickpea. However, there are limited studies on its molecular mechanisms. To address this issue, we conducted integrated transcriptomic and bioinformatics analyses for species-specific genes and acyl-lipid-, starch-, and carbon metabolism-related genes. Among seven expressional patterns of soybean-specific genes, four were highly expressed at the middle- and late oil accumulation stages; these genes significantly enriched fatty acid synthesis and carbon metabolism, and along with common acetyl CoA carboxylase (ACCase) highly expressed at soybean middle seed development stage, common starch-degrading enzyme beta-amylase-5 (BAM5) was highly expressed at soybean early seed development stage and oil synthesis-related genes ACCase, KAS, KAR, ACP, and long-chain acyl-CoA synthetase (LACS) were co-expressed with WRI1, which may result in high seed oil content and low seed starch content in soybean. The common ADP-glucose pyrophosphorylase (AGPase) was highly expressed at chickpea middle seed development stage, along with more starch biosynthesis genes co-expressed with four-transcription-factor homologous genes in chickpea than in soybean, and the common WRI1 was not co-expressed with oil synthesis genes in chickpea, which may result in high seed starch content and low seed oil content in chickpea. The above results may be used to improve chickpea seed oil content in two ways. One is to edit *CaWRI1* to co-express with oil synthesis-related genes, which may increase carbon metabolites flowing to oil synthesis, and another is to increase the expression levels of miRNA159 and miRNA319 to inhibit the expression of MYB33, which may downregulate starch synthesis-related genes, making more carbon metabolites flow into oil synthesis. Our study will provide a basis for future breeding efforts to increase the oil content of chickpea seeds.

Keywords: lipid synthesis, starch synthesis, gene regulation network, soybean, chickpea

INTRODUCTION

Legumes are an indispensable part of the human diet. They provide one-third of dietary protein and are also an important source of animal feed and edible oil (Zhu et al., 2005). As we know, there are cold- and warm-season legumes. The former, such as chickpea and pea, are rich in starch, while the latter, such as soybean and peanut, are rich in oil and protein (Rathi et al., 2016). Although seed oil and starch content in soybean and chickpea depend on the supply of carbon in plants (Weselake et al., 2009), there are limited studies on the molecular mechanisms for the differences in seed oil and starch content of the two crops.

In recent decades, efforts have been made to dissect the molecular mechanism of seed oil and starch synthesis. In seed oil synthesis, there are four main steps: fatty acid *de novo* synthesis, acyl elongation and editing, triacylglycerol (TAG) assembly, and oil droplet formation. The expression levels of the relevant genes in these steps affect seed oil content, e.g., acetyl CoA carboxylase (ACCase) in rapeseed (Roesler et al., 1997) and potato (Klaus et al., 2004), β -Ketoacyl-[acyl carrier protein] synthase (KASI) in *Arabidopsis*, rice, and soybean (Wu and Xue, 2010; Ding et al., 2015; Dobbels et al., 2017), diacylglycerol acyltransferase (DGAT) in peanut, *Arabidopsis*, and oil palm (Saha et al., 2006; Hernández et al., 2012; Rosli et al., 2018), and oleosins (OLE) in *Arabidopsis* (Shimada and Hara-Nishimura, 2010). Seed starch synthesis begins with the generation of adenosine diphosphate (ADP) glucose *via* the enzyme ADP-glucose pyrophosphorylase (AGPase). AGPase (Tang et al., 2016), soluble starch synthase (SS) (Cuesta-Sejio et al., 2016), granule-bound starch synthases (GBSS) (Denyer et al., 1999), starch branching enzyme (SBE) (Lu et al., 2015), and isoamylase (ISA) (Bustos et al., 2004) are key enzymes in starch synthesis. Transcription factors (TFs) can simultaneously regulate the expression of multiple genes. Key TFs in the regulation of lipid biosynthesis include WRI1 (Ma et al., 2015), LEC1 (Tan et al., 2011), LEC2 (Baud et al., 2007), FUS3 (Wang et al., 2007), and ABI3 (Crowe et al., 2000). TFs in the biosynthesis of starch include SUSH3A2 (Sun et al., 2003), OsbZIP58 (Wang et al., 2013), OsMADS6 and OsMADS29 (Nayar et al., 2013), ABI4 (Hu et al., 2012), and ZmMYB138 and ZmMYB115 (Hu et al., 2021). The above enzymes and TFs may help us to dissect the molecular mechanisms for the differences in seed oil and starch content of soybean and chickpea.

In recent years, some studies have focused on the molecular mechanism of complex trait formation. To dissect the photosynthetic mechanism of C_4 and C_3 crops, for example, Wang et al. (2014) compared their genomes and transcriptomes and found that higher expression levels of orthologous genes related to photosynthetic development at the base of rice leaves compared to that in maize, *cis* element (RGCGR; R = A/G) that is only present in promoters of maize, and 118 TFs that contributed to the expression levels of C_4 photosynthetic cell-type-specific genes, may result in the photosynthesis differences between C_4 and C_3 crops. To dissect the molecular mechanism of seed starch content between adzuki bean and soybean, Yang et al. (2015) performed comparative genomic and transcriptome analyses and found no significant variation of starch biosynthesis genes in

the two genomes, but the increased expressional levels of these genes at the mature seed stage in adzuki bean and the decreased expressional levels of these genes at the seed filling and mature stages in soybean may result in the difference of seed starch content. To dissect the molecular mechanism of seed oil content between soybean and rapeseed, Zhang et al. (2018) compared genomes, transcriptomes, and miRNA regulation and found that the inhibition of *BnPEPC1* by bna-miR169, high expression levels of rape-specific genes encoding acetyl-coenzyme β -CT, and biotin carboxyl carrier protein-1 (BCCP1) subunits at the seed development stage, along with the expansion of rapeseed lipid storage-related genes and the contraction of lipid degradation-related genes, may result in high seed oil content in rapeseed. However, the molecular mechanisms for the differences of seed oil and starch content in soybean and chickpea are unclear.

To dissect the molecular mechanism for the differences of seed oil and starch content in soybean and chickpea, integrated transcriptomic and bioinformatics analyses for species-specific genes and acyl-lipid-, starch-, and carbon metabolism-related genes, along with miRNA, TF, and co-expression network analysis, were carried out in this study. As a result, it was found that ACCase, BAM5, AGPase, and WRI1 may be related to the above difference. In addition, we discussed how to improve seed oil content in chickpeas. The results provide an explanation for the differences in seed oil and starch content of the two crops.

MATERIALS AND METHODS

Genomic and Proteomic Data

The protein-coding sequences of all the genes in *Arabidopsis*, soybean, peanut, pea, and chickpea were downloaded from <https://www.arabidopsis.org/index.jsp> (TAIR10), <https://www.soybase.org/> (Wm82.a2.v1), https://legumeinfo.org/data/public/Arachis_hypogaea (arahy.Tifrunner.CCJH), https://legumeinfo.org/data/public/Pisum_sativum (Cameor.gnm1.ann1.7SZR), and <https://phytozome.jgi.doe.gov/pz/portal.html> (Carietinum_492_v1.0), respectively. If there were multiple transcripts for a gene, the longest one was selected.

Transcriptome Data

Transcriptome datasets in soybean, chickpea, peanut, and pea were downloaded from <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE42871> (GSE42871; Jones and Vodkin, 2013), <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79719> and <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE79720> (GSE79719 & GSE79720; Garg et al., 2017), https://peanutbase.org/gene_expression/ atlas (Clevenger et al., 2016 and Alves-Carvalho et al., 2015), respectively. The soybean datasets included seven developmental stages: whole seed 4 days after fertilization (DAF) (R2); whole seed 12–14 DAF (R3); whole seed 22–24 DAF (R4); whole seed 5–6 mg in weight (R5); cotyledons 5–6 mg in weight, seed coats 5–6 mg in weight, cotyledons 100–200 mg in weight (R6); cotyledons 400–500 mg in weight (R7); and dry whole seed (R8). The chickpea datasets included S1–S7 seed development stages and leaf. The stages S1–S3 were the embryo development stages (embryogenesis), stages S4 and S5 were early and mid-maturation stages, respectively

(grain filling), and stages S6 and S7 were late mature stages (seed desiccation). Based on the definition of soybean vegetative and reproductive growth (Ritchie et al., 1982), whole seed 12–14 DAF (R3), whole seed 22–24 DAF (R4), whole seed 5–6 mg in weight (R5), cotyledons 400–500 mg in weight (R7), and dry whole seed (R8) in soybean were almost consistent with the stages S2, S3, S4, S5, and S7 in chickpea, respectively. These stages were uniformly defined as the stages t1, t2, t3, t4, and t5 in this study, respectively. Thus, the five stages were selected to conduct expressional trend analysis. The peanut datasets included Pattee 5 seed, Pattee 6 seed, Pattee 7 seed, Pattee 8 seed, and Pattee 10 seed. The pea datasets included Stem_BC_LN, Pods_C_LN, seeds_5dai, seeds_5dai_mut, Seeds_12dap, Seeds_12dap_mut, RootSys_A_HN, RootSys_A_LN, Leaf_B_LN, and Flower_B_LN. The gene expression levels [reads per kilobase per million mapped (RPKM) reads] were normalized and quantified by the DESeq package in Bioconductor. The package DESeq provides methods to test for differential expression by estimating variance-mean dependence in count data from high-throughput sequencing assays and test for differential expression based on a model.

Zhang et al. (2018) described that the relative expression level was the ratio of the expression level of each gene to the average expression level of all genes in this species.

Identification of Common and Specific Genes in Soybean and Chickpea

In this study, all the non-redundant whole-genome protein sequences in *Arabidopsis*, soybean, peanut, pea, and chickpea were clustered by the software package OrthoFinder 2.0 (Emms and Kelly, 2019) to identify orthologous gene families. All the genes of soybean and chickpea were analyzed by MCscan (Tang et al., 2008) to identify homologous collinearity gene fragments.

Co-expressional Network Analysis for Specific Genes in Soybean and Chickpea

The expressional levels of soybean- and chickpea-specific genes at the above seven different seed development stages and tissues were analyzed by R package WGCNA v1.68 to construct co-expressional networks. When calculating the scale-free topological overlap matrix (TOM), the parameters “unsigned” and “bicor” were used. The optimal soft thresholds in soybean and chickpea were calculated by the function “pickSoftThreshold,” and their thresholds were 10. The minimum module size was set to 30, and the “merge cut height” parameter of the module was set to 0.3 (Langfelder and Horvath, 2008). The KEGG enrichment analysis for the genes in the above co-expressional networks was conducted by KOBAS (<http://kobas.cbi.pku.edu.cn/index.php>, version 2.0; Xie et al., 2011).

Cluster Analysis for Expressional Levels of Common Genes in Soybean and Chickpea

The expressional levels of soybean and chickpea common genes at the t1–t5 stages were used to perform cluster analysis, as described in Zhang et al. (2018), using Short Time-series Expression Miner (STEM, <http://www.cs.cmu.edu/~jernst/stem/>) (Ernst and Bar-Joseph, 2006) with the following parameters: log normalize a time-series vector of gene expression values ($v_0, v_1, v_2, \dots, v_n$) to $[0, \log_2(v_1/v_0), \log_2(v_2/v_0), \dots, \log_2(v_n/v_0)]$, minimum absolute expression change 2, $-p$ 0.05.

Copy Number Analysis of Gene Families Related to Carbon, Lipid, and Starch Metabolism

Copy Number Analysis of Gene Families Related to Carbon, Lipid, and Starch Metabolism

One hundred and thirty-five genes related to fatty acid synthesis, TAG synthesis, lipid droplet assembly, and storage were downloaded from ARALIP (<http://aralip.plantbiology.msu.edu/>; Li-Beisson et al., 2013; Zhang et al., 2018). Thirty-seven starch metabolism-related genes in *Arabidopsis* were mined from Schwarte et al. (2015). Using the method of Troncoso-Ponce et al. (2011), 238 genes related to carbon metabolism were obtained. The copy number difference analysis was carried out for the above 410 genes (Supplementary Table 4). Meanwhile, the 410 genes were compared with the ones in categories three and four of the STEM analysis to mine common genes related to carbon, lipid, and starch metabolism.

Identification of microRNA, TFs, and Their Target Genes

The microRNAs in soybean and chickpea were downloaded from miRbase (21st edition, <http://www.mirbase.org/>) and Jain et al. (2014), respectively. The online software psRNATarget (<http://plantgrn.noble.org/psRNATarget/>) (Dai and Zhao, 2011) was used to identify miRNA targets with default parameters except for the Expectation (e), which was set at 3. The TFs were downloaded from PlantTFDB 3.0 (<http://planttfdb.cbi.pku.edu.cn/>) (Jin et al., 2017). The expressional levels of 88 gene families obtained from copy number analysis and the target TFs of microRNA were used to calculate the Pearson correlation coefficient.

Gene Family Phylogenetic Tree Construction

The gene families of *WRII* in soybean, chickpea, *Arabidopsis*, peanut, and pea were used to construct a gene phylogenetic tree using the neighbor-joining (NJ) method, implemented by software MEGA7 (<http://www.megasoftware.net/>, Kumar et al., 2016); the bootstrap test was repeated 1,000 times, using Poisson distribution model and “pairwise deletion” option. The iTOL (<http://itol.embl.de/>) was used to visualize and post-beautify the final result tree file.

Conserved Domain Analysis of Gene and Protein Sequences

The online tool MEME (<http://meme-suite.org/tools/meme>, v4.11.2) was used to predict the conserved domains of key protein-coding genes (Bailey and Elkan, 1994). The Clustalw program in the MEGA7 software was used to perform a multiple sequence alignment of amino acid sequences of the *WRII* homologous gene family in *Arabidopsis*, soybean, chickpea, peanut, and pea with default parameters. The

GeneDoc software (<https://github.com/karlnicholas/GeneDoc>, Nicholas et al., 1997) was used to compare and beautify the results.

RESULTS

Identification of Common and Specific Genes in Soybean and Chickpea

The collinearity analysis tool MCscan was used to analyze the collinearity of soybean and chickpea genomes. As a result, there were 31,519 and 1,443 repeated gene pairs in the soybean and chickpea genomes, respectively. The latter was far fewer than the former, which may be due to the soybean-specific tetraploid (Wang et al., 2017). In soybean, 472, 305, and 152 blocks had more than 10, 20, and 50 collinear genes, which contained 29,101 (52.8%), 26,689 (48.4%), and 21,172 (39.5%) collinear genes, respectively. In chickpea, 45, 21, and 5 blocks had more than 10, 20, and 50 collinear genes, which contained 1,214 (4.3%), 830 (2.9%), and 380 (1.3%) collinear genes, respectively (**Supplementary Table 2**).

All the 223,609 protein-coding genes in soybean, chickpea, *Arabidopsis*, peanut, and pea were clustered by OrthoFinder to identify homologous gene families in soybean and chickpea. As a result, these genes were clustered into 26,575 orthologous groups (OGs) (**Supplementary Table 1**), and each OG represents a gene family. Among these OGs, there were 335 OGs with both soybean single-copy and chickpea multi-copy genes, 8,675 OGs with both soybean multi-copy and chickpea single-copy genes, 3,261 OGs with soybean and chickpea multi-copy genes, and 2,396 OGs with soybean and chickpea single-copy genes, which are regarded as common genes, while 18,657 (33.29%) were defined as soybean-specific paralogous gene clusters without chickpea genes and only 8,010 (28.33%) were chickpea-specific genes (**Supplementary Table 3**; Zhang et al., 2018).

Co-expression and KEGG Analyses of Specific Genes in Soybean or Chickpea

The expression levels of 18,657 soybean-specific genes at seven different seed development stages and tissues (Jones and Vodkin, 2013) were analyzed by R package WGCNA v1.68 to obtain different co-expression modules. As a result, seven co-expression modules with different expression patterns were identified. The genes in each module were used to conduct KEGG enrichment analysis. Results showed that four modules, blue (3,318 genes), black (900 genes), green (563 genes), and light green (250 genes), significantly enriched multiple lipid metabolism pathways ($P < 0.05$; **Figure 1B**). The genes in the blue module were highly expressed between R2 and R5 seed development stages, and significantly enriched not only the carbohydrate metabolism pathways of pyruvate metabolism (gmx00620), pentose phosphate pathway (gmx00030), carbon metabolism (gmx01200), and glycolysis (gmx00010), but also the glycerolipid metabolism

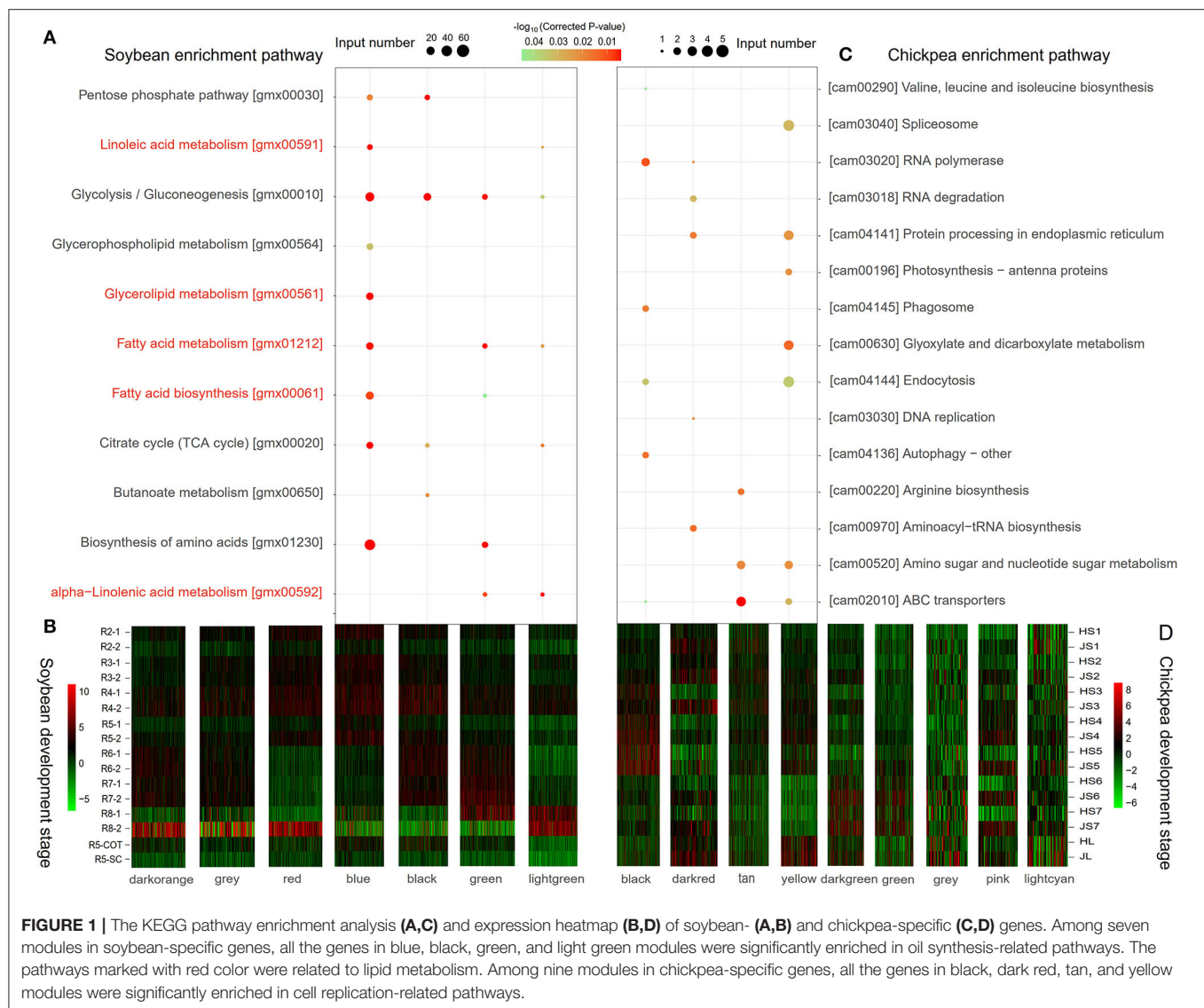
pathway (gmx00561), glycerophospholipid metabolism (gmx00564), linoleic acid metabolism (gmx00591), and other metabolic pathways. Meanwhile, these genes enriched the photosynthesis-antenna protein (gmx00196) and carbon fixation (gmx00710) metabolic pathways in photosynthetic organisms. Therefore, the blue module is considered to be a co-expression module for oil accumulation during soybean seed storage material accumulation (**Figure 1A**).

The expression levels of 8,010 chickpea-specific genes at seven different seed development stages and tissues (Garg et al., 2017) were also analyzed. As a result, nine co-expression modules with different expression patterns were identified. The genes in the black (573 genes), tan (190 genes), yellow (335 genes), and dark red (101 genes) modules were highly expressed at S3–S5 stages, and they enriched cell replication and synthesis-related metabolic pathways. This may be related to the rapid expansion of the number of seed cells (**Figures 1C,D**).

Expressional Trends of Common Genes Between Soybean and Chickpea

To obtain the expression trends of soybean and chickpea common genes, the standardized expression levels of 37,387 soybean and 20,259 chickpea common genes at different seed development stages (**Supplementary Figure 1**) were used to conduct cluster analysis with the *k*-means algorithm via the STEM software. As a result, these common genes were clustered into six clusters. Among these clusters, the expression of all the genes in Cluster3–Cluster5 was upregulated during the t2–t4 stages, while the expression of all the genes in Cluster1 was downregulated during the t2–t4 stages.

All the genes in each cluster were used to conduct KEGG metabolic pathway enrichment analysis. As a result, all the soybean genes in Cluster3 and Cluster4 enriched pentose phosphate pathway (gmx00030), fructose and mannose metabolism (gmx00051), galactose metabolism (gmx00052), fatty acid biosynthesis (gmx00061), fatty acid degradation (gmx00071), starch and sucrose metabolism (gmx00500), pyruvate metabolism (gmx00620), glyoxylic acid and dicarboxylic acid metabolism (gmx00630), carbon fixation in photosynthetic organisms (gmx00710), fatty acid metabolism (gmx01212), and other metabolic pathways (**Supplementary Figure 2A**). Meanwhile, all the chickpea genes in Cluster4 mainly enriched fatty acid biosynthesis (cam00061), fatty acid metabolism (cam01212), glycolysis/gluconeogenesis (cam00010), citrate cycle (TCA cycle) (cam00020), carbon fixation in photosynthetic organisms (cam00710), glycerophospholipids metabolism (cam00564), starch and sucrose metabolism (cam00500), glyoxylate and dicarboxylate metabolism (cam00630), and other pathways (**Supplementary Figure 2B**). Clearly, the above enriched metabolic pathways were found to be related to seed lipid and starch synthesis, and the expression of these genes was upregulated during the rapid accumulation stages of these substances, indicating the existence of fatty acids and starch in soybean and chickpea.



Identification of Candidate Oil- and Starch Biosynthesis-Related Genes in Common Genes

Four hundred and ten carbon-, lipid-, and starch metabolism-related genes in *Arabidopsis thaliana* (Supplementary Table 4) were used as the seed to search for gene families in 26,575 OGs. As a result, 374 common gene families were mined. Among these gene families, six had a >2 copy number ratio of soybean to chickpea, while eight had a copy number ratio of <1. The former includes GRF2 during photosynthesis, CT- α and Acyl carrier protein (ACP) involved in the fatty acid synthesis, and CDS, PECT, and DGAT participating in TAG synthesis, while the latter includes STP7 involved in carbohydrate transport, PKp- β involved in glycolysis, Beta-hydroxyacyl-acyl carrier protein dehydratase (HAD) and β -Ketoacyl-acyl carrier protein synthase II (KAS II) participating in the fatty acid synthesis, glycerol-3-phosphate acyltransferase (GPAT9), LPAAT3, and FAD2 participating in TAG synthesis, and Pyrophosphorylase

(PPA) involved in starch metabolism (Supplementary Figure 3). Meanwhile, the above 410 genes were traversed and the genes in Cluster3 and Cluster4 were searched to obtain homologous gene families. As a result, 158 soybean genes and 121 chickpea genes were found to be homologous with the above 410 genes.

By merging the above 279 upregulated genes and the above 14 gene families with copy number difference, 184 soybean and 135 chickpea genes in 97 OGs were obtained (Supplementary Table 5), which are candidate oil- and starch biosynthesis-related common genes in soybean and chickpea, and some genes have been confirmed by molecular biological experiments (Table 1).

Expression Trends of Candidate Oil and Starch Synthesis Genes

To compare the expressional profile difference of each gene in the above 97 OGs at t1–t5 stages in soybean and chickpea, the relative expression level was used in this study (Supplementary Table 6).

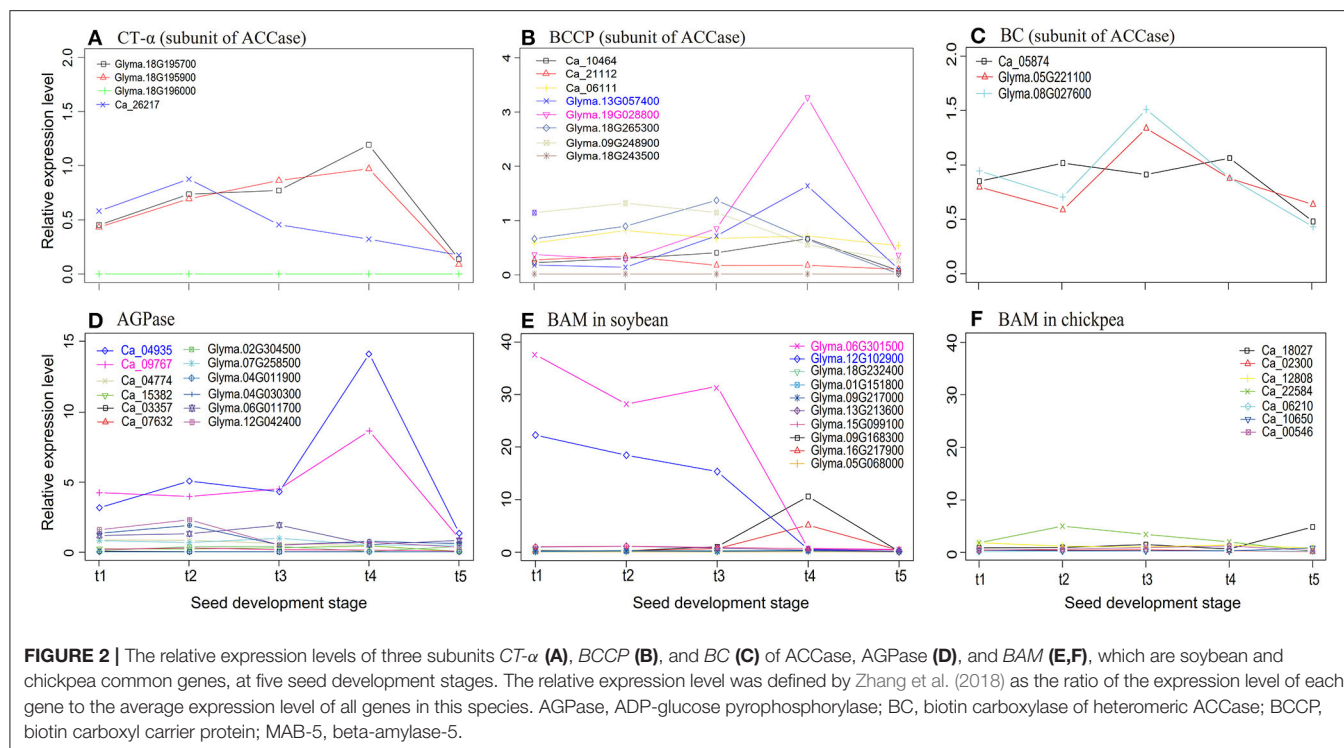
TABLE 1 | The confirmed candidate genes for the differences of soybean and chickpea seed oil and starch content in previous studies.

Biological process	Known genes in <i>Arabidopsis</i>			Comparative genomics analysis		References
	<i>Arabidopsis</i>	Gene	Functional annotation	Homologs in soybean	Homologs in chickpea	
Plastid fatty acid synthesis from pyruvate	AT5G16390	BCCP1	Biotin Carboxyl Carrier Protein of Heteromeric ACCase	Glyma.18G243500 Glyma.09G248900	Ca_06111	Klaus et al., 2004
	AT5G35360	BC	Biotin Carboxylase of Heteromeric ACCase	Glyma.08G027600 Glyma.05G221100	Ca_05874	Roesler et al., 1997
	AT5G46290	KASI	Ketoacyl-ACP Synthase I	Glyma.08G084300	Ca_05157	Yang et al., 2016
	AT1G74960	KASII	Ketoacyl-ACP Synthase II	Glyma.17G047000	Ca_03125 Ca_19284	Aslan et al., 2015
	AT1G62640	KASIII	Ketoacyl-ACP Synthase III	Glyma.18G211400 Glyma.09G277400	Ca_02927	Choi et al., 2000
	AT1G77590	LACS	Long-Chain Acyl-CoA Synthetase	Glyma.06G112900 Glyma.13G079900 Glyma.20G060100	Ca_12668 Ca_13132	Zhao et al., 2010
TAG synthesis	AT5G60620	GPAT9	Glycerol-3-Phosphate Acyltransferase	Glyma.09G119200	Ca_17353 Ca_25130	Singer et al., 2016
	AT1G51260	LPAAT3	1-Acylglycerol-3-Phosphate Acyltransferase	Glyma.15G034100	Ca_02646 Ca_05840	Liu et al., 2015
	AT2G19450	DGAT1	Acyl-CoA: Diacylglycerol Acyltransferase	Glyma.13G106100 Glyma.09G065300 Glyma.17G053300	Ca_03178	Routaboul et al., 1999
	AT3G51520	DGAT2	Acyl-CoA: Diacylglycerol Acyltransferase	Glyma.01G156000	Ca_10695	Gao et al., 2021
	AT5G13640	PDAT1	Phospholipid: Diacylglycerol Acyltransferase	Glyma.07G036400 Glyma.13G108100 Glyma.16G005800	Ca_03160 Ca_10036	Fan et al., 2013
	AT1G12640	LPCAT	1-acylglycerol-3-phosphocholine Acyltransferase	Glyma.17G131500	Ca_08736	Bates et al., 2012
Starch metabolism	AT5G48300	ApS1	ADP-glucose pyrophosphorylase	Glyma.02G304500	Ca_07632 Ca_09767	Li et al., 2011
	AT5G19220	ApL1	Glucose-1-phosphate adenyllyltransferase	Glyma.07G258500	Ca_03357	Tuncel et al., 2014
	AT5G24300	SSI	Starch synthase	Glyma.04G235200	Ca_03985	Nakamura, 2002
	AT3G01180	SSII	Starch synthase	Glyma.13G062700 Glyma.19G022900	Ca_10512	Craig et al., 1998
	AT1G11720	SSIV	Starch synthase	Glyma.05G127800	Ca_05169	Szydlowski et al., 2009
	AT4G18240	SSIII	Starch synthase	Glyma.13G204700	Ca_00636	Gao et al., 1998
	AT1G32900	GBSS	Granule-bound starch synthase	Glyma.20G218100 Glyma.07G049900	Ca_22418	Ceballos et al., 2007
	AT2G39930	ISA1	Isoamylase1	Glyma.08G028400	Ca_05882	Delatte et al., 2005
	AT3G20440	BE2	Isoamylase3	Glyma.18G092600	Ca_20526	Dumez et al., 2006
	AT4G15210	BAM5	Beta-amylase	Glyma.12G102900 Glyma.06G301500	Ca_22584	Andriotis et al., 2010

In soybean, most starch synthesis-related genes were highly expressed at the early seed development stage, while most oil synthesis-related genes were highly expressed in the middle seed development stage, which is also the rapid oil accumulation stage. In chickpea, most starch synthesis-related genes were highly expressed at the middle seed development stage.

Based on the studies in Klaus et al. (2004), Tang et al. (2016), and Andriotis et al. (2010), common genes *ACCase*, *AGPase*, and *BAM5* were important genes in seed oil and starch metabolism.

Thus, we compared the expressional profile differences of these genes in soybean and chickpea. First, the relative expression levels of genes encoding subunits α -CT, BCCP, and BC of ACCase at rapid oil accumulation stages were higher in soybean than in the chickpea (**Figures 2A–C**). Because ACCase catalyzes the first step of the fatty acid synthesis, which is the formation of malonyl-CoA from acetyl-CoA (Klaus et al., 2004), ACCase may make more carbon metabolites flow into the fatty acid synthesis in soybean seeds. Then, the relative expression levels of genes encoding



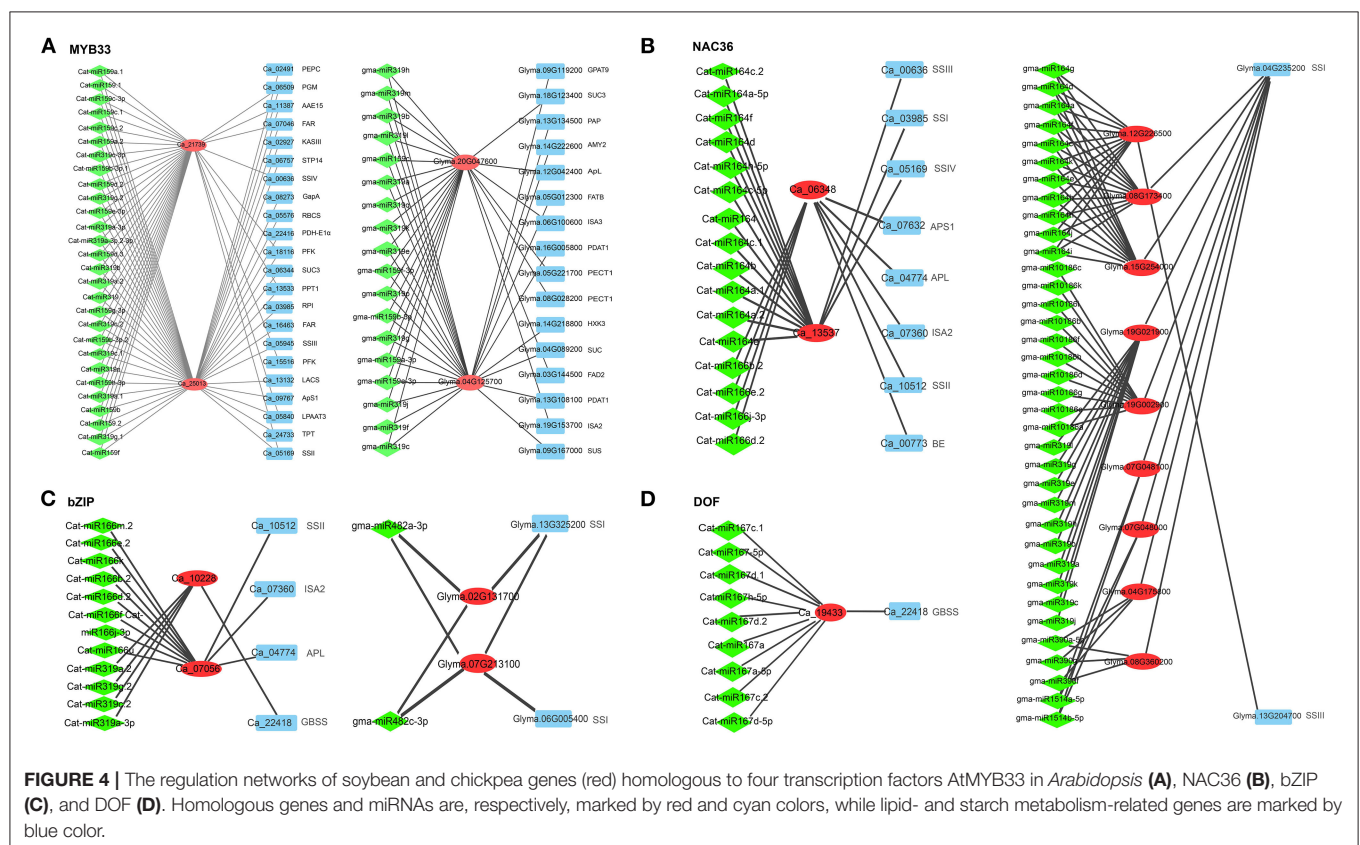
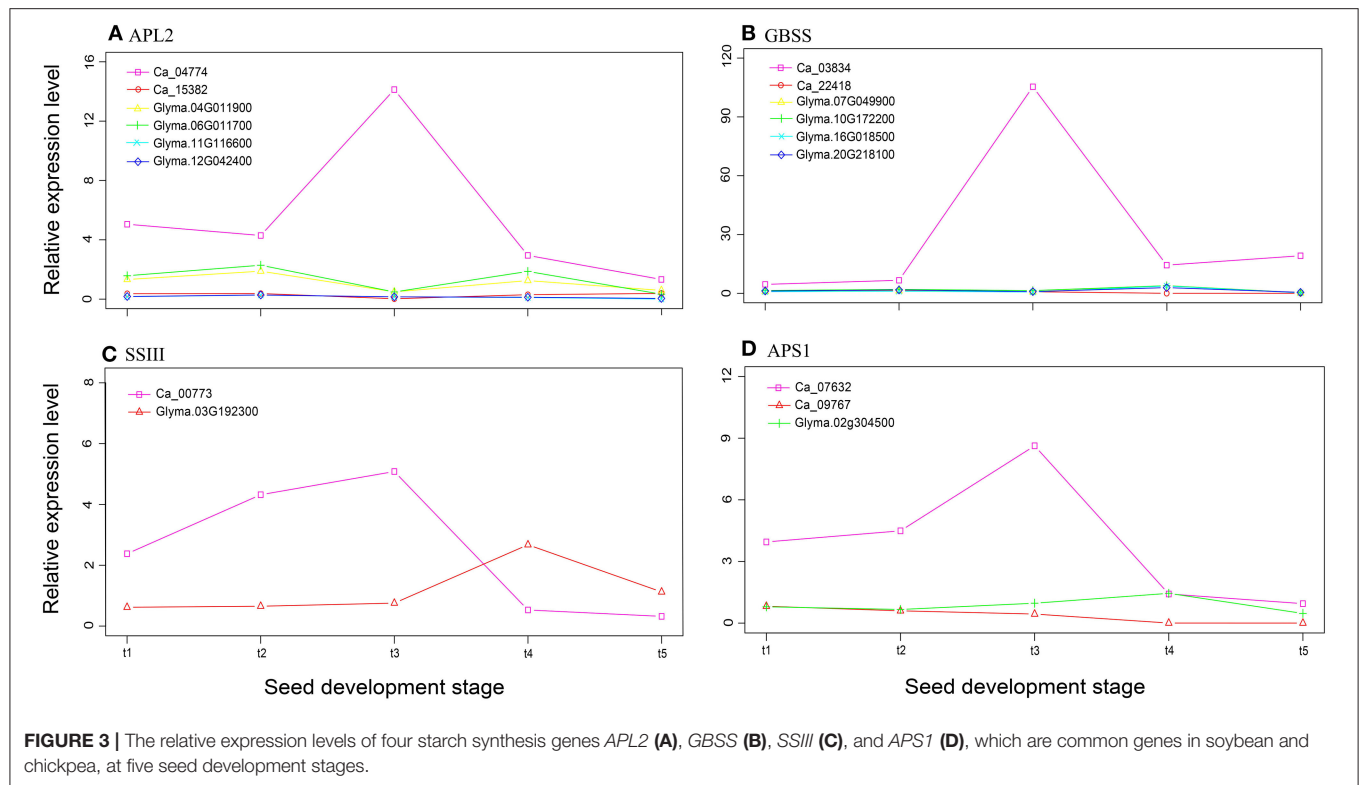
subunits ADP-glucose pyrophosphorylase (APS) and glucose-1-phosphate adenylyltransferase (APL) of *AGPase* at the t4 stage were much higher in chickpea than in soybean (Figure 2D).

Because *AGPase* is the first key regulatory enzyme and rate-limiting enzyme in starch synthesis (Tang et al., 2016), *AGPase* may play a significant role in chickpea starch synthesis. Finally, the relative expression level of *BAM5* at the early seed development stage was much higher in soybean than in chickpea (Figures 2E,F). Because *BAM5* catalyzes the production of maltose from linear glucans (Andriotis et al., 2010), it may play a major role in plant starch hydrolysis. This may result in low starch content in soybean seeds. Therefore, we infer that *ACCase*, *AGPase*, and *BAM5* were more likely to be candidate key genes that affect the differences of seed oil and starch content in soybean and chickpea.

Using the starch biosynthesis genes in rice (17 genes; Ohdan et al., 2005; Tian et al., 2009) and *Arabidopsis* (13 genes; Schwarte et al., 2015) as bait, we performed an ortholog search in the soybean and chickpea genomes. In total, fewer starch biosynthesis genes were found in chickpea than in soybean (33 vs. 45; Supplementary Table 7), but there was no significant difference between these two genomes in the exact Fisher test ($P = 1.00$). The relative expression levels of these starch biosynthesis genes at seven seed development stages were used to compare their differences. We found that the average relative expression level in chickpea was significantly higher than that of soybean, e.g., *Ca_03834*, *Ca_04774*, *Ca_07632*, and *Ca_00773* have a higher expressional level at stage t3 (Figure 3), although more starch biosynthesis genes were detected in the soybean genome (Supplementary Table 7). The above results are similar to those in Yang et al. (2015).

Comparison of Regulatory Networks of miRNAs, TFs, and Their Target Genes in Soybean and Chickpea

The target genes of 639 soybean and 440 chickpea mature miRNAs were predicted using the online software psRNATarget. As a result, the target genes in 2,303 soybean and 939 chickpea miRNA-target gene pairs were TFs, while the target genes in 90 soybean and 30 chickpea miRNA-target gene pairs were oil- or starch synthesis-related genes. In detail, in soybean *gma-miR2111* targeted *SSIII* (*Glyma.13G204700*), *gma-miR5784*- and *gma-miR4347* targeted *APS1* (*Glyma.02G304500*), and *gma-miR10413* targeted *PGM2* (*Glyma.20G018000*); *gma-miR530*, *gma-miR160*, and *gma-miR319* were predicted to regulate the expression of the *ABI4*, *ARF*, and *MYB* gene families, respectively, while *gma-miR156* and *gma-miR159* were predicted to regulate the expression of the *SBP* gene family (Supplementary Table 8). In chickpea, *Cat-miR156* and *Cat-miR157* inhibited the expression of *phospholipid: diacylglycerol acyltransferase-2* (*PDAT2*; *Ca_24053*), while *Cat-miR171* and *Cat-miR5565* inhibited the expression of *FATB* (*Ca_06618*). *Cat-miR166* and *Cat-miR319* were predicted to regulate the expression of *bZIP* in chickpea, while *gma-miR482* was predicted to regulate the expression of *bZIP* in soybean. *Cat-miR164* and *gma-miR164* were predicted to regulate the expression of *NAC36* in chickpea and soybean, respectively. *Cat-miR167* was predicted to regulate the expression of *DOF* in chickpea, while *gma-miR172* was predicted to regulate the expression of *DOF* in soybean (Supplementary Table 9; Figure 4). *PDAT* transfers acyl moiety from PC to DAG to



form TAG, which plays an important role in TAG synthesis. Cat-miR160, Cat-miR5674, and Cat-miR156 were predicted to regulate the expression of *ARF*, *GRAS*, and *SBP* gene families, respectively, while Cat-miR156, Cat-miR159, and Cat-miR319 were predicted to regulate the expression of the *MYB* gene family (Supplementary Table 8).

The Regulatory Networks of miRNAs, TFs, and Their Target Genes

As described by Hu et al. (2021), genes involved in a biological network are regulated by the same TF and co-expressed with each other. In the above 2,303 soybean and 939 chickpea miRNA-target gene pairs, there were 457 soybean TFs and 113 chickpea TFs. These TFs were merged with 11 oil synthesis-related TFs and 6 starch synthesis-related homologous gene families in soybean and chickpea. After removing unexpressed TFs, 282 soybean TFs and 93 chickpea TFs were identified. Soybean transcriptomic datasets of 282 TFs and 184 genes at seven different seed development stages and tissues (Jones and Vodkin, 2013) were used to calculate their correlation coefficients (r), while chickpea transcriptomic datasets of 93 TFs and 135 genes at seven different seed development stages and tissues (Garg et al., 2017) were used to calculate their correlation coefficients. With the thresholds of $|r| \geq 0.8$ and $P \leq 0.05$, a total of 1,905 soybean TF-gene pairs and 590 chickpea TF-gene pairs were found to be significant.

In the regulatory network, miR156, miR159, miR160, and miR319 families in chickpea and soybean commonly regulated *MYB* and *SBP*. We found that two chickpea genes and two soybean genes, homologous to *AtMYB33*, had different co-expressed genes, e.g., *Ca_21739* was co-expressed with *PFK5*, *FBA*, *LPAAT3*, *SSIV*, *SSII*, and *AMY2*, and *Ca_25013* was co-expressed with *STP7*, *PFK5*, *FBA*, *GapA*, *TPT*, *PDH-E1 α* , *KASIII*, *LPAAT3*, *LPEAT2*, *LACS*, *APS1*, *SSI*, and *SSIV*, while *Glyma.04G125700* was co-expressed with *AMY2*, *SUC*, *SUS*, *GPAT9*, *FATB*, and *PDAT1*, and *Glyma.20G047600* was co-expressed with *SUC3*, *FAD2*, *PECT1*, *APL*, *ISA2*, and *ISA3* (Figure 4A). The four genes were all regulated by miRNA159 and miRNA319. It is speculated that miRNA159 and miRNA319 may regulate the expression of oil- and starch synthesis-related genes via regulating *MYB* gene families in crops. Two chickpea genes, *Ca_10228* and *Ca_07056*, homologous to *bZIP*, had more co-expressed starch biosynthesis genes than two soybean genes, *Glyma.02G131700* and *Glyma.07G213100*, which are homologous to *bZIP*; in detail, *bZIP* was co-expressed with starch biosynthesis genes *APL*, *ISA2*, *SSII*, and *GBSS* in chickpea but only with starch biosynthesis gene *SSI* in soybean (Dong et al., 2019; Figure 4C). Two chickpea genes *Ca_06348* and *Ca_13537*, homologous to *NAC36*, had more co-expressed starch biosynthesis genes than three soybean genes, *Glyma.08G173400*, *Glyma.12G226500*, and *Glyma.15G254000*, which are homologous to *NAC36*; in detail, *NAC36* was co-expressed with starch biosynthesis genes *APL*, *APS1*, *BE*, *ISA2*, *SSI*, *SSII*, *SSIII*, and *SSIV* in chickpea but only with starch biosynthesis genes *SSI* and *SSIII* in soybean (Zhang et al., 2014; Figure 4B). *DOF* was co-expressed with one starch biosynthesis gene *Ca_22418* (*GBSS*) in chickpea but

with no starch biosynthesis genes in soybean (Qi et al., 2017; Supplementary Table 9; Figure 4D). It is speculated that miRNA164, miRNA166, miRNA167, miRNA 319, and miRNA482 may regulate the expression of starch synthesis-related genes via regulating *bZIP*, *NAC36*, and *DOF* gene families in chickpea and/or soybean. This is a possible reason that seed starch content is higher in chickpea than in soybean.

GmWRI1 (*Glyma.08G227700* and *Glyma.15G221600*) were co-expressed with lipid synthesis genes *a-CT*, *BC*, *BCCP2*, *KASII*, *KAR*, *LACS*, *FATA*, *PKP- α* , *PKP- β* , *ACP*, and 1-acylglycerol-3-phosphocholine acyltransferase (*LPCAT*) in soybean, while no lipid synthesis genes in chickpea were found to be co-expressed with *CaWRI1* (*Ca_15711*). Surprisingly, no miRNA was found to regulate *WRI1*. The fact that no lipid synthesis genes co-expressed with *WRI1* may be an important reason for low seed oil content in chickpea. Thus, we conducted further analysis for the *WRI1* gene family.

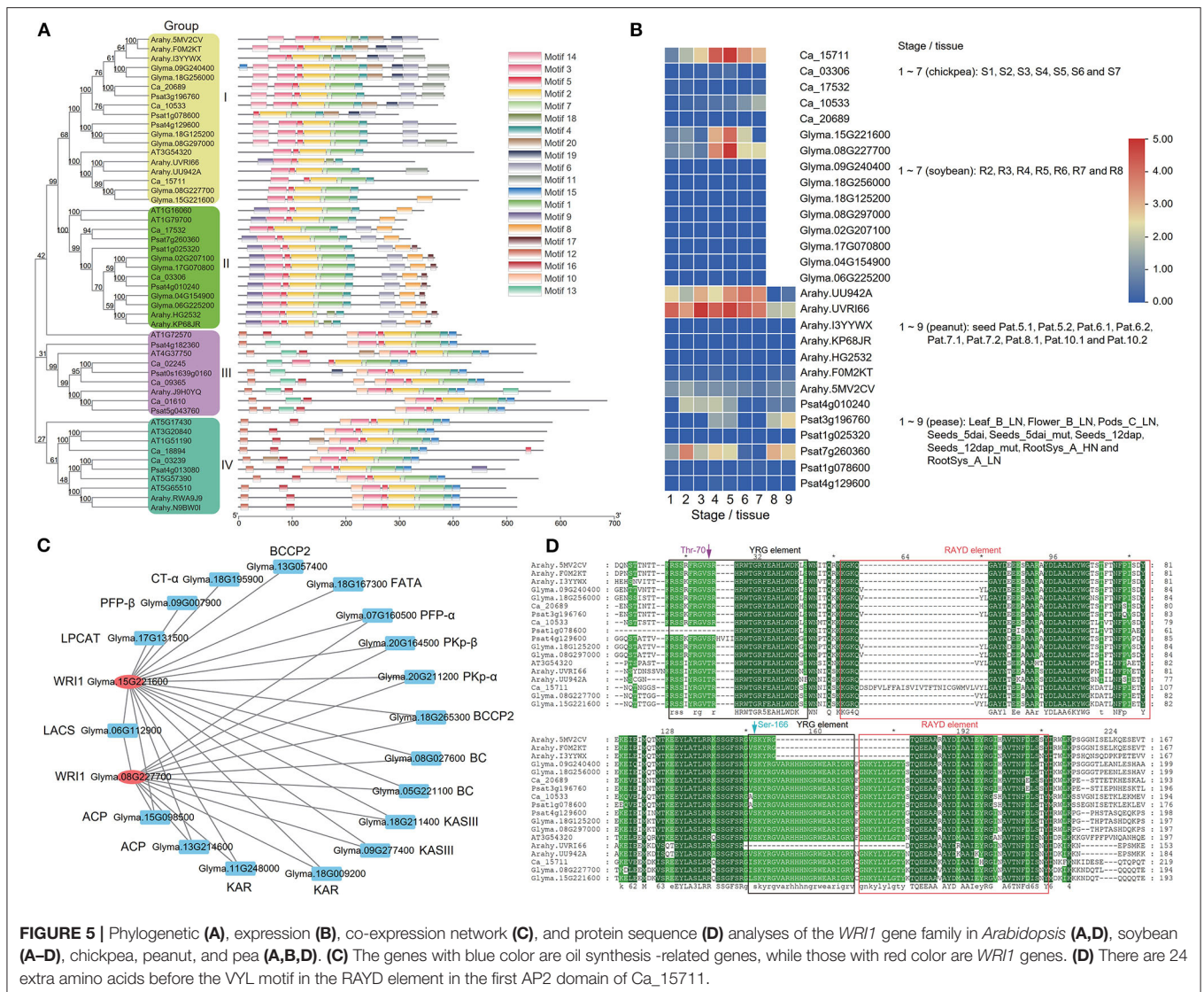
Phylogenetic Analysis of *WRI1* Gene Family

Construction of Phylogenetic Tree of *WRI1* Gene Family

As a member of the AP2 subfamily of the AP2/EREBP TF family, *WRI1* plays an important role in the regulation of seed oil synthesis (Tang et al., 2019). To investigate the evolution of the *WRI1* gene family, all the *WRI1* homologs in soybean, chickpea, *Arabidopsis*, peanut, and pea were searched by BLASTP using the *Arabidopsis WRI1* amino-acid sequence as the query sequence. As a result, 50 sequences with $E < 1e-10$ were downloaded for these species and were used to construct a phylogenetic tree using a NJ method via MEGA7. In the evolutionary tree, there were four *WRI* sub-families, which are named Groups I–IV. Among these subfamilies, *AtWRI1* was in Group I, while *AtWRI4* was in Group II (Figure 5A). To identify their conserved motifs, all the protein sequences were analyzed by the software MEME. The results are shown in Figure 5A. Although the sequence motifs were very similar in Groups I and II, some small differences existed, e.g., *Psat1g078600* in Group I lacked some front part sequences, and *Ca_15711* had a large gap between the first and second conserved motifs. Meanwhile, different sequence motifs in Groups III and IV were observed.

Expressional Patterns of *WRI1* Subfamilies

To understand the functions of all the *WRI* genes in Groups I and II in soybean, chickpea, peanut, and pea, we compared the expressional patterns of all the *WRI* genes at different tissues or seed developmental stages. In Group I, *Ca_15711*, *Glyma.15G221600*, *Glyma.08G227700*, and *Arahy.UU942A* had low expression levels at the early seed development stage and high expression levels at lipid synthesis stages (Figure 5B), while *Psat4g010240* and *Psat4g010240* expressed at the early seed development stage and did not express at later stages. In addition, only *Ca_10533* had low expression levels at late seed development stages, while no other soybean and chickpea *WRI* genes in Group I were expressed. In Group II, most *WRI* genes in soybean, peanut, and chickpea had low or no expression levels, while only



Psat7g260360 in pea was expressed at the early seed development stage. The results are consistent with those of Lee et al. (2009), in which the mRNAs of *WRI3* and *WRI4* were almost undetectable.

Correlation Analysis of *WRI1* Genes With Lipid Synthesis Genes

To compare the correlation of *WRI1* genes in Groups I and II with lipid synthesis genes in warm- (peanut and soybean) and cool-season (chickpea and pea) legumes, we calculated correlation coefficients of *WRI1* genes in peanut and pea and lipid synthesis genes. Using the thresholds of $|r| \geq 0.8$ and $P \leq 0.05$, we identified a total of 64 TF-gene pairs. Among these pairs, there were 31, 31, 2, and 0 TF-gene pairs in peanut, soybean, pea, and chickpea, respectively, which were all in Group I rather than in Group II (Table 2; Figure 5C).

Protein Sequence Analysis of All the *WRI1* Genes

To understand the structural variation of all the *WRI1* genes, we performed multiple alignments for the protein

sequence of all the *WRI1* genes in Group I. The results are shown in Figure 5D. As described by Okamuro et al. (1997), *WRI1* has two AP2/EREb domains, and in each domain, YRG and RAYD elements form the AP2 DNA-binding domain. We found that all the *WRI1* genes, except *Arahy.UVRI66* and *Psat1g078600* included two YRG elements and two RAYD elements. The second YRG element was more conserved than the first one. The tyrosine (Y) in the first YRG motif is conserved in *AT3G54320*, *Arahy.UVRI66*, *Arahy.UU942A*, *Ca_15711*, *Glyma.08G227700*, and *Glyma.5G221600*, while the alanine (A) residue in the first RAYD motif and the arginine (R) residue in the second RAYD motif are not conserved. In the first AP2 domain of *Psat1g078600*, some sequences in the YRG element were lacking, while in the first AP2 domain of *Ca_15711*, 24 extra amino acids before the VYL motif in the RAYD element were observed. The above sequence variation may affect the functions of the *WRI1* gene in different species.

TABLE 2 | Co-expressional gene pairs of *WRI1* with oil synthesis genes in peanut, soybean, and pea.

Species	Transcription factor	Co-expressional gene	Gene annotation	r	P-value	Species	Transcription factor	Co-expressional gene	Gene annotation	r	P-value
<i>Arachis hypogea</i>	<i>Arahy.UU942A</i>	<i>Arahy.AG8095</i>	<i>BCCP2</i>	0.9576	4.97E-05	<i>Glycine max</i>	<i>Glyma.08G227700</i>	<i>Glyma.20G164500</i>	<i>PKp-β</i>	0.8896	3.98E-06
	<i>Arahy.UU942A</i>	<i>Arahy.7G97IV</i>	<i>SAD</i>	0.9546	6.28E-05		<i>Glyma.08G227700</i>	<i>Glyma.20G211200</i>	<i>PKp-α</i>	0.9200	4.55E-07
	<i>Arahy.UU942A</i>	<i>Arahy.QR9FOJ</i>	<i>FATA</i>	0.8895	1.32E-03		<i>Glyma.08G227700</i>	<i>Glyma.18G265300</i>	<i>BCCP2</i>	0.9125	8.36E-07
	<i>Arahy.UU942A</i>	<i>Arahy.YI95N7</i>	<i>FATA</i>	0.8436	4.26E-03		<i>Glyma.08G227700</i>	<i>Glyma.08G027600</i>	<i>BC</i>	0.9215	3.99E-07
	<i>Arahy.UU942A</i>	<i>Arahy.0G9KRE</i>	<i>GPAT9</i>	0.8946	1.13E-03		<i>Glyma.08G227700</i>	<i>Glyma.05G221100</i>	<i>BC</i>	0.9347	1.15E-07
	<i>Arahy.UVRI66</i>	<i>Arahy.45ZUWH</i>	<i>CT-α</i>	0.8506	3.66E-03		<i>Glyma.08G227700</i>	<i>Glyma.18G211400</i>	<i>KASIII</i>	0.9013	1.88E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.KBA53Y</i>	<i>BCCP2</i>	0.9054	7.80E-04		<i>Glyma.08G227700</i>	<i>Glyma.09G277400</i>	<i>KASIII</i>	0.8776	7.93E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.1S754L</i>	<i>BC</i>	0.8433	4.28E-03		<i>Glyma.08G227700</i>	<i>Glyma.18G009200</i>	<i>KAR</i>	0.8806	6.74E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.H4HC75</i>	<i>BC</i>	0.8813	1.69E-03		<i>Glyma.08G227700</i>	<i>Glyma.11G248000</i>	<i>KAR</i>	0.8677	1.33E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.GCU9ZU</i>	<i>KASI</i>	0.8432	4.30E-03		<i>Glyma.08G227700</i>	<i>Glyma.13G214600</i>	<i>ACP</i>	0.8525	2.73E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.X916ZF</i>	<i>KASI</i>	0.8408	4.52E-03		<i>Glyma.08G227700</i>	<i>Glyma.15G098500</i>	<i>ACP</i>	0.8290	7.21E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.39SMC6</i>	<i>KASII</i>	0.8931	1.18E-03		<i>Glyma.08G227700</i>	<i>Glyma.06G112900</i>	<i>LACS</i>	0.9678	8.81E-10
	<i>Arahy.UVRI66</i>	<i>Arahy.F6FJSK</i>	<i>KASII</i>	0.8359	5.00E-03		<i>Glyma.15G221600</i>	<i>Glyma.09G007900</i>	<i>PFP-β</i>	0.8323	6.34E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.40PBNN</i>	<i>KAR</i>	0.8709	2.24E-03		<i>Glyma.15G221600</i>	<i>Glyma.07G160500</i>	<i>PFP-α</i>	0.8730	1.02E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.ICU8XT</i>	<i>KAR</i>	0.8390	4.69E-03		<i>Glyma.15G221600</i>	<i>Glyma.20G164500</i>	<i>PKp-β</i>	0.8375	5.17E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.JC4ZCG</i>	<i>HAD</i>	0.9485	9.73E-05		<i>Glyma.15G221600</i>	<i>Glyma.20G211200</i>	<i>PKp-α</i>	0.8923	3.38E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.KP3Y1K</i>	<i>HAD</i>	0.9457	1.17E-04		<i>Glyma.15G221600</i>	<i>Glyma.18G195900</i>	<i>CT-α</i>	0.8028	1.81E-04
	<i>Arahy.UVRI66</i>	<i>Arahy.EA8RPB</i>	<i>ACP</i>	0.9227	3.92E-04		<i>Glyma.15G221600</i>	<i>Glyma.13G057400</i>	<i>BCCP2</i>	0.8434	4.04E-05
	<i>Arahy.UVRI66</i>	<i>Arahy.FL5QND</i>	<i>ACP</i>	0.8600	2.94E-03		<i>Glyma.15G221600</i>	<i>Glyma.18G265300</i>	<i>BCCP2</i>	0.9521	1.36E-08
	<i>Arahy.UVRI66</i>	<i>Arahy.Z9XA8H</i>	<i>ACP</i>	0.9527	7.22E-05		<i>Glyma.15G221600</i>	<i>Glyma.08G027600</i>	<i>BC</i>	0.9386	7.52E-08
	<i>Arahy.UVRI66</i>	<i>Arahy.YI95N7</i>	<i>FATA</i>	0.8744	2.04E-03		<i>Glyma.15G221600</i>	<i>Glyma.05G221100</i>	<i>BC</i>	0.9523	1.33E-08
	<i>Arahy.UVRI66</i>	<i>Arahy.HDYS9D</i>	<i>LACS</i>	0.9433	1.35E-04		<i>Glyma.15G221600</i>	<i>Glyma.18G211400</i>	<i>KASIII</i>	0.8761	8.63E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.L14VXE</i>	<i>ACC2</i>	0.8115	7.92E-03		<i>Glyma.15G221600</i>	<i>Glyma.09G277400</i>	<i>KASIII</i>	0.8922	3.39E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.0G9KRE</i>	<i>GPAT9</i>	0.8691	2.35E-03		<i>Glyma.15G221600</i>	<i>Glyma.18G009200</i>	<i>KAR</i>	0.9060	1.35E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.F7JGYB</i>	<i>DGAT3</i>	0.8782	1.84E-03		<i>Glyma.15G221600</i>	<i>Glyma.11G248000</i>	<i>KAR</i>	0.8948	2.88E-06
	<i>Arahy.UVRI66</i>	<i>Arahy.3LN7S1</i>	<i>LACS</i>	0.8336	5.23E-03		<i>Glyma.15G221600</i>	<i>Glyma.13G214600</i>	<i>ACP</i>	0.9018	1.82E-06
	<i>Arahy.5MV2CV</i>	<i>Arahy.M024AA</i>	<i>HAD</i>	0.8437	4.25E-03		<i>Glyma.15G221600</i>	<i>Glyma.15G098500</i>	<i>ACP</i>	0.8711	1.12E-05
	<i>Arahy.5MV2CV</i>	<i>Arahy.Q99U9Q</i>	<i>HAD</i>	0.8316	5.45E-03		<i>Glyma.15G221600</i>	<i>Glyma.18G167300</i>	<i>FATA</i>	0.8245	8.52E-05
	<i>Arahy.5MV2CV</i>	<i>Arahy.I1811Q</i>	<i>LACS</i>	0.8139	7.59E-03		<i>Glyma.15G221600</i>	<i>Glyma.06G112900</i>	<i>LACS</i>	0.9830	1.07E-11
	<i>Arahy.5MV2CV</i>	<i>Arahy.F78IM4</i>	<i>PDAT1</i>	0.8067	8.61E-03		<i>Glyma.15G221600</i>	<i>Glyma.17G131500</i>	<i>LPCAT</i>	0.8269	7.79E-05
	<i>Arahy.5MV2CV</i>	<i>Arahy.WLZ7Z3</i>	<i>FAD3</i>	0.8433	4.29E-03	<i>Pisum sativum</i>	<i>Psat3g196760</i>	<i>Psat7g144640</i>	<i>KAR</i>	0.9660	2.53E-04
	<i>Glycine max</i>	<i>Glyma.08G227700</i>	<i>Glyma.07G160500</i>	<i>PFP-α</i>	0.8140	1.25E-04	<i>Psat3g196760</i>	<i>Psat7g130880</i>	<i>PDAT</i>	0.9916	2.04E-02

DISCUSSION

Molecular Mechanism of High Oil and Low Starch Content in Soybean Seed

As described above, the soybean-specific genes in four of seven modules were highly expressed at the middle and late oil accumulation stages and significantly enriched in fatty acid, glycerolipid, and linoleic acid metabolisms. Among these genes, main lipid synthesis-related genes include *Glyma.19G147400* (Lakhssassi et al., 2017), *Glyma.07G207200* (Deng et al., 2020), *Glyma.12G146700* (Carther et al., 2019), and *Glyma.06G058500* (Yang et al., 2021), which have been confirmed by molecular biology experiments. In detail, *Glyma.19G147400* is responsible for the conversion of oleic acid to linoleic acid, *Glyma.07G207200* catalyzes the synthesis of monounsaturated oleic acid or palmitoleic acid in plastids, *Glyma.12G146700* transforms the diacylglycerol into phosphatidic acid, and *Glyma.06G058500* is involved in long-chain fatty acids synthesis. Meanwhile, around stable loci for soybean seed oil content in Yao et al. (2020), *Glyma.05G226800*, *Glyma.05G227400*, *Glyma.08G102100*, *Glyma.10G063600*, *Glyma.11G148900*, *Glyma.11G190600*, *Glyma.13G112700*, *Glyma.13G148100*, *Glyma.14G150100*, *Glyma.15G051100*, and *Glyma.20G088000* were found to be soybean-specific genes in the above four modules. Thus, the above soybean-specific and highly expressed genes may result in high seed oil content.

Three soybean copies (*Glyma.13G057400*, *Glyma.18G265300*, and *Glyma.19G028800*) of common gene *BCCP2* in soybean and chickpea were highly expressed during t3–t4 stages; in particular, *Glyma.19G028800* was rapidly upregulated and reached a peak during t3–t4 stages. As described in Konishi and Sasaki (1994), *BCCP2* is an important subunit of *ACCase*. Overexpression of *ACCase* can significantly increase fatty acid content in crop seed (Roesler et al., 1997; Klaus et al., 2004), while antisense expression of *BCCP2* transcript in developing *Arabidopsis* seeds resulted in an average 38% reduction in *BCCP2* protein, further leading to an average 9% reduction in fatty acid content in mature seeds (Thelen and Ohlrogge, 2002). It should be noted that *Glyma.13G057400* and *Glyma.19G028800* (*BCCP2*) in soybean had significantly higher expression than *Ca_21112* and *Ca_10464* (*BCCP2*) in chickpea. Thus, the high expression genes *Glyma.13G057400* and *Glyma.19G028800* in soybean may result in its high seed oil content.

In this study, *WRI1* was found to be co-expressed with oil synthesis-related genes *ACCase*, *KAS*, *KAR*, *ACP*, and *LACS* in soybean and peanut. This may result in high seed oil content in soybean. The evidence is as follows. Some target genes of *AtWRI1* are involved in *Arabidopsis* glycolysis and fatty acid synthesis, while their mRNAs are accumulated in a synergistic manner and regulated by *WRI1* (Baud et al., 2009). Twenty-eight transcripts, such as *KASIII*, *KAR*, and *LACS*, were found to have increased expression levels in the transgenic plants of *GmWRI1*, which can specifically bind the 500 bp upstream sequences of ATG codon of the above genes and effectively activate their transcription (Chen et al., 2018).

As described in Scheidig et al. (2002), *BAM* degrades starch and amylopectin to form maltose. In the study of Andriotis et al. (2010), seed starch content was 1.5–2.5 times higher in *Arabidopsis thaliana bam4* and *bam1234* mutants than in wild-type embryos, and the starch still existed in mature embryos, while seed starch content in *amy3*, *isa3*, and *phs1* mutants did not increase, indicating the necessity of *BAM* in embryo starch degradation. In this study, two *BAM5* copies, *Glyma.06G301500* and *Glyma.12G102900*, in soybean had more than 6 times expression levels at the early seed development stage than one chickpea *BAM5* gene *Ca_22584*. This indicates that seed starch, produced at the early seed development stage, maybe degraded by *Glyma.06G301500* and *Glyma.12G102900*. This results in low seed starch content in soybean. *NAC36*-, *bZIP*-, and *DOF*-homologous genes in chickpea have more co-expressed starch biosynthesis genes than those in soybean. This may result in higher seed starch content in chickpea than in soybean.

The above results are summarized in **Figures 4, 6**.

Molecular Mechanism of High Starch and Low Oil Content in Chickpea Seed

Although *CaWRI1* in chickpea has a similar expression pattern as its homologous genes *Glyma.08G227700* and *Glyma.5G221600* in soybean, no chickpea oil synthesis-related genes were co-expressed with *CaWRI1*. The possible reason is 24 extra amino acids of the *RAYD* element of the first AP2 domain of *CaWRI1* before the *VYL* motif. In previous studies, mutant seeds with T-DNA inserted in the 5th exon of *AtWRI1* decreased seed oil content 72% in *Arabidopsis* than in its wild type (An et al., 2017), while *wri1* mutant could not convert glucose and sucrose into precursors of fatty acid biosynthesis during seed development, resulting in the decrease of seed oil content (Focks and Benning, 1998). Thus, it is speculated that the structural variation of *CaWRI1* leads to the failure to regulate glycolysis and lipid synthesis-related genes, which may be an important reason for low seed oil content in chickpea.

As described above, *AGPase* had more than 7 times relative expression level in chickpea than in soybean. Its high expression is likely to catalyze more Glc-1-P to ADP glucose (Crevillén et al., 2005). As evidenced in rice (Tuncel et al., 2014) and potato (Hajirezaei et al., 2003), *AGPase* could effectively control starch flux. In molecular biology, the 100-grain weight of maize seeds with overexpressing *AGPase* was increased by 15%, and seed starch content was increased by 74% compared to its wild type (Li et al., 2011), while the weight of wheat seeds with overexpressing cytoplasmic *AGPase* large subunit gene was increased by 9.1%, and seed starch content was increased by 9.6% compared to its wild type (Kang et al., 2013). Thus, high expression of *AGPase* is more likely to result in higher seed starch content in chickpea.

The above results are summarized in **Figure 6**.

Yang et al. (2015) performed the comparative genomic and transcriptome analyses of seed starch and oil content between

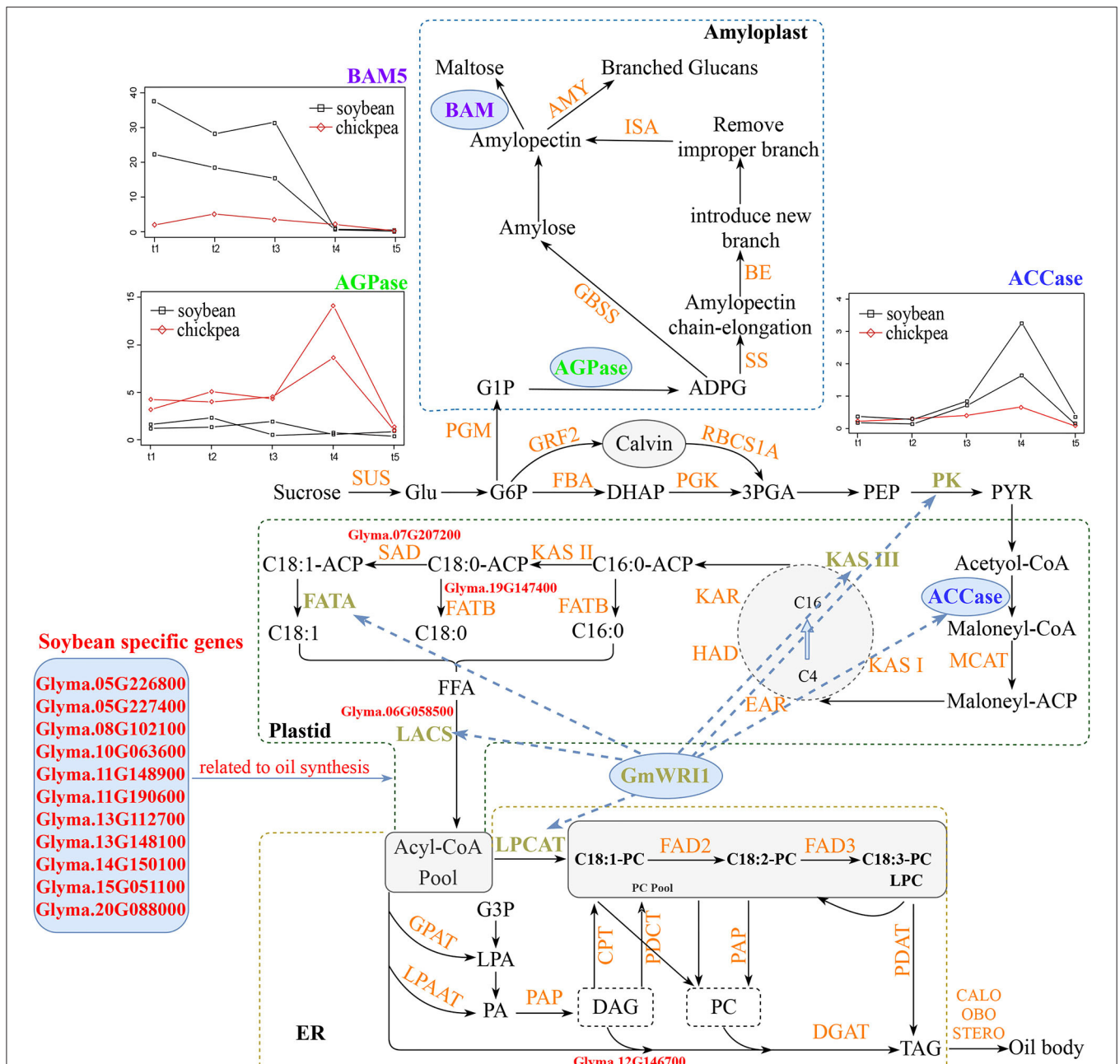


FIGURE 6 | Molecular mechanisms for the difference in seed oil and starch content between soybean and chickpea. In soybean, soybean-specific fatty acid synthesis genes (red), common highly expressed ACCase (blue), and common oil synthesis-related genes ACCase, KAS, KAR, ACP, and LACS co-expressed with *WRI1* (brown) are responsible for high seed oil content, while common starch degrading enzyme BAM5 (purple) highly expressed at soybean early seed development stage is responsible for low seed starch content. In chickpea, common AGPase (green) highly expressed at chickpea middle seed development stage is responsible for high seed starch content, while no oil synthesis genes co-expressed with common *WRI1* is responsible for low seed oil content. LACS, long-chain acyl-CoA synthetase; AGPase, ADP-glucose pyrophosphorylase; BAM-5, beta-amylase-5; ACCase, acetyl CoA carboxylase.

adzuki bean and soybean in the genome sequencing project of adzuki bean (*Vigna angularis*). In this project, they found no significant variation of starch biosynthesis genes in the two genomes, meanwhile, they checked the expression levels of five key genes that encode the proteins involved in the

conversion from starch to oil synthesis. Although Yang et al. (2015) did not compare starch synthesis-related genes, key TFs, miRNAs, and their regulation networks, oil- and starch synthesis-related genes, TFs, miRNAs, and their regulatory networks in soybean and chickpea were investigated in this study to obtain

the molecular mechanisms for the differences in seed oil and starch content.

Breeding by Design for High Seed Oil Content in Chickpea

As pointed out by Dragičević et al. (2015), increasing seed oil content in chickpea can reduce seed phytic acid content and improve its digestibility. However, there is low seed oil content in the present chickpea cultivars. To address this issue, it is necessary to improve seed oil content in chickpea. Based on the above results, two genetic improvement methods are proposed.

First, deleting the 24 extra amino acids of the RAYD element in Ca_15711 (CaWRI1) may make the edited CaWRI1 co-express with oil synthesis-related genes to increase seed oil content in chickpea. In this study, we found that the relative expression level of Ca_15711 (CaWRI1) at stage S3 in chickpea is higher than those of soybean genes *Glyma.15G221600* and *Glyma.08G227700*, homologous to Ca_15711, at stage R4 (Figure 5B). Shen et al. (2010) had provided transgenic evidence. In detail, seed oil content in maize was significantly higher in transgenic *ZmWRI1* endosperm (0.81%) than in null endosperm (0.47%), but seed starch content was reduced approximately 60%, although the expression of *ZmWRI1* does not affect protein content in embryo and grain yield, indicating that *ZmWRI1* may enhance oil biosynthesis by reducing carbon metabolites in starch biosynthesis.

Second, increasing the expression of miRNA159 and miRNA319 in chickpea may decrease the expression of MYB33, which downregulates starch synthesis-related enzymes AGPase, SS, and ISA, making more carbon metabolites flow into fatty acid synthesis. Hu et al. (2021) provided experimental evidence. In detail, Zma-miR159k-3p could negatively regulate the expression of *ZmMYB138* and *ZmMYB115*, homologous to *AtMYB33*, in maize endosperm, while *ZmMYB138* and *ZmMYB115* may be positively correlated with amylopectin and amylose content in

maize endosperm, respectively. Thus, Zma-miR159k-3p may indirectly regulate starch synthesis in maize endosperm by regulating the expression of *ZmMYBs*.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

Y-MZ conceived and supervised the study. KC, Y-FP, L-ML, and H-QZ carried out the experimental works and analyzed the data. KC, Y-FP, and Y-MZ wrote and revised the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (32070557 and 31871242) and the Huazhong Agricultural University Scientific and Technological Self-innovation Foundation (2014RC020).

ACKNOWLEDGMENTS

The authors thank Mr. Hanwen Zhang (Hence Education Ltd., Vancouver, Canada; hywenzhang@henceedu.com) for improving the language within the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.743680/full#supplementary-material>

REFERENCES

- Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochet, A. L., Jacquin, F., et al. (2015). Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J.* 84, 1–19. doi: 10.1111/tpj.12967
- An, D., Kim, H., Ju, S., Go, Y. S., Kim, H. U., and Suh, M. C. (2017). Expression of *Camelina* WRINKLED1 isoforms rescue the seed phenotype of the *Arabidopsis* wr1 mutant and increase the triacylglycerol content in tobacco leaves. *Front. Plant Sci.* 8:34. doi: 10.3389/fpls.2017.00034
- Andriotis, V. M., Pike, M. J., Kular, B., Rawsthorne, S., and Smith, A. M. (2010). Starch turnover in developing oilseed embryos. *New Phytol.* 187, 791–804. doi: 10.1111/j.1469-8137.2010.03311.x
- Aslan, S., Hofvander, P., Dutta, P., Sitbon, F., and Sun, C. (2015). Transient silencing of the KASII genes is feasible in *Nicotiana benthamiana* for metabolic engineering of wax ester composition. *Sci. Rep.* 5:11213. doi: 10.1038/srep11213
- Bailey, T. L., and Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 2, 28–36.
- Bates, P. D., Fathi, A., Snapp, A. R., Carlsson, A. S., Browse, J., and Lu, C. (2012). Acyl editing and headgroup exchange are the major mechanisms that direct polyunsaturated fatty acid flux into triacylglycerols. *Plant Physiol.* 160, 1530–1539. doi: 10.1104/pp.112.204438
- Baud, S., Mendoza, M. S., To, A., Harscoët, E., Lepiniec, L., and Dubreucq, B. (2007). WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. *Plant J.* 50, 825–838. doi: 10.1111/j.1365-313X.2007.03092.x
- Baud, S., Wuillème, S., To, A., Rochat, C., and Lepiniec, L. (2009). Role of WRINKLED1 in the transcriptional regulation of glycolytic and fatty acid biosynthetic genes in *Arabidopsis*. *Plant J.* 60, 933–947. doi: 10.1111/j.1365-313X.2009.04011.x
- Bustos, R., Fahy, B., Hylton, C. M., Seale, R., Nebane, N. M., Edwards, A., et al. (2004). Starch granule initiation is controlled by a heteromultimeric isoamylase in potato tubers. *Proc. Natl. Acad. Sci. U.S.A.* 101, 2215–2220. doi: 10.1073/pnas.0305920101
- Carther, K. F. I., Keteouli, T., Ye, N., Yang, Y. H., Wang, N., Dong, Y. Y., et al. (2019). Comprehensive genomic analysis and expression profiling of diacylglycerol kinase (DGK) gene family in soybean (*Glycine max*) under abiotic stresses. *Int. J. Mol. Sci.* 20:1361. doi: 10.3390/ijms20061361
- Ceballos, H., Sánchez, T., Morante, N., Fregene, M., Dufour, D., Smith, A. M., et al. (2007). Discovery of an amylose-free starch mutant in cassava (*Manihot esculenta* Crantz). *J. Agric. Food Chem.* 55, 7469–7476. doi: 10.1021/jf070633y

- Chen, M., Lin, J. Y., Hur, J., Pelletier, J. M., Baden, R., Pellegrini, M., et al. (2018). Seed genome hypomethylated regions are enriched in transcription factor genes. *Proc. Natl. Acad. Sci. U.S.A.* 115, 8315–8322. doi: 10.1073/pnas.1811017115
- Choi, K. H., Heath, R. J., and Rock, C. O. (2000). β -ketoacyl-acyl carrier protein synthase III (FabH) is a determining factor in branched-chain fatty acid biosynthesis. *J. Bacteriol.* 182, 365–370. doi: 10.1128/JB.182.2.365-370.2000
- Clevenger, J., Chu, Y., Scheffler, B., and Ozias-Akins, P. (2016). A developmental transcriptome map for allotetraploid *Arachis hypogaea*. *Front. Plant Sci.* 7:1446. doi: 10.3389/fpls.2016.01446
- Craig, J., Lloyd, J. R., Tomlinson, K., Barber, L., Edwards, A., Wang, T. L., et al. (1998). Mutations in the gene encoding starch synthase II profoundly alter amylopectin structure in pea embryos. *Plant Cell* 10, 413–426. doi: 10.1105/tpc.10.3.413
- Crevillén, P., Ventriglia, T., Pinto, F., Orea, A., Mérida, A., and Romero, J. M., et al. (2005). Differential pattern of expression and sugar regulation of *Arabidopsis thaliana* ADP-glucose pyrophosphorylase-encoding genes. *J. Biol. Chem.* 280, 8143–8149. doi: 10.1074/jbc.M411713200
- Crowe, A. J., Abenes, M., Plant, A., and Moloney, M. M., et al. (2000). The seed-specific transactivator, ABI3, induces oleosin gene expression. *Plant Sci.* 151, 171–181. doi: 10.1016/S0168-9452(99)00214-9
- Cuesta-Sejio, J. A., Nielsen, M. M., Ruzanski, C., Kruciewicz, K., Beeren, S. R., Rydhal, M. G., et al. (2016). *In vitro* biochemical characterization of all barley endosperm starch synthases. *Front. Plant Sci.* 6:1265. doi: 10.3389/fpls.2015.01265
- Dai, X., and Zhao, P. X. (2011). psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 39, 155–159. doi: 10.1093/nar/gkr319
- Delatte, T., Trevisan, M., Parker, M. L., and Zeeman, S. C. (2005). *Arabidopsis* mutants Atisal and Atisa2 have identical phenotypes and lack the same multimeric isoamylase, which influences the branch point distribution of amylopectin during starch synthesis. *Plant J.* 41, 815–830. doi: 10.1111/j.1365-313X.2005.02348.x
- Deng, M., Liu, B., Wang, Z., Xue, J., Zhang, H., and Li, R. (2020). Identification and functional analysis of soybean stearyl-ACP Δ^9 desaturase (*GmSAD*) gene family. *Chin. J. Biotechnol.* 36, 716–731. doi: 10.13345/j.cjb.190550
- Denyer, K., Waite, D., Motawia, S., Möller, B. L., and Smith, A. M. (1999). Granule-bound starch synthase I in isolated starch granules elongates maltotriose oligosaccharides processively. *Biochem. J.* 340, 183–191. doi: 10.1042/bj3400183
- Ding, W., Lin, L., Zhang, B., Xiang, X., Wu, J., Pan, Z., et al. (2015). OsKASI, a β -ketoacyl-[acyl carrier protein] synthase I, is involved in root development in rice (*Oryza sativa* L.). *Planta* 242, 203–213. doi: 10.1007/s00425-015-2296-2
- Dobbels, A. A., Michno, J. M., Campbell, B. W., Virdi, K. S., Stec, A. O., Muehlbauer, G. J., et al. (2017). An induced chromosomal translocation in soybean disrupts a KASI ortholog and is associated with a high-sucrose and low-oil seed phenotype. *G3* 7, 1215–1223. doi: 10.1534/g3.116.038596
- Dong, Q., Xu, Q., Kong, J., Peng, X., Zhou, W., Chen, L., et al. (2019). Overexpression of *ZmbZIP22* gene alters endosperm starch content and composition in maize and rice. *Plant Sci.* 283, 407–415. doi: 10.1016/j.plantsci.2019.03.001
- Dragičević, V., Kratovalieva, S., Dumanović, Z., Dimov, Z., and Kravić, N. (2015). Variations in level of oil, protein, and some antioxidants in chickpea and peanut seeds. *Chem. Biol. Technol. Agric.* 2:2. doi: 10.1186/s40538-015-0031-7
- Dumez, S., Wattedled, F., Dauville, D., Delvalle, D., Planchot, V., Ball, S. G., et al. (2006). Mutants of *Arabidopsis* lacking starch branching enzyme II substitute plastidial starch synthesis by cytoplasmic maltose accumulation. *Plant Cell* 18, 2694–2709. doi: 10.1105/tpc.105.037671
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:238. doi: 10.1186/s13059-019-1832-y
- Ernst, J., and Bar-Joseph, Z. (2006). STEM: a tool for the analysis of short time series gene expression data. *BMC Bioinformatics* 7:191. doi: 10.1186/1471-2105-7-191
- Fan, J., Yan, C., Zhang, X., and Xu, C. (2013). Dual role for phospholipid:diacylglycerol acyltransferase: enhancing fatty acid synthesis and diverting fatty acids from membrane lipids to triacylglycerol in *Arabidopsis* leaves. *Plant Cell* 25, 3506–3518. doi: 10.1105/tpc.113.117358
- Focks, N., and Benning, C. (1998). wrinkled1: a novel, low-seed-oil mutant of *Arabidopsis* with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* 118, 91–101. doi: 10.1104/pp.118.1.91
- Gao, M., Wanat, J., Stinard, P. S., James, M. G., and Myers, A. M. (1998). Characterization of dull1, a maize gene coding for a novel starch synthase. *Plant Cell* 10, 399–412. doi: 10.1105/tpc.10.3.399
- Gao, Y., Sun, Y., Gao, H., Chen, Y., Wang, X., Xue, J., et al. (2021). Ectopic overexpression of a type-II DGAT (CeDGAT2-2) derived from oil-rich tuber of *Cyperus esculentus* enhances accumulation of oil and oleic acid in tobacco leaves. *Biotechnol. Biofuels* 14:76. doi: 10.1186/s13068-021-01990-2
- Garg, R., Singh, V. K., Rajkumar, M. S., Kumar, V., and Jain, M., et al. (2017). Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *Plant J.* 91, 1088–1107. doi: 10.1111/tpj.13621
- Hajirezaei, M. R., Börnke, F., Peisker, M., Takahata, Y., Lerchl, J., Kirakosyan, A., et al. (2003). Decreased sucrose content triggers starch breakdown and respiration in stored potato tubers (*Solanum tuberosum*). *J. Exp. Bot.* 54, 477–488. doi: 10.1093/jxb/erg040
- Hernández, M. L., Whitehead, L., He, Z., Gazda, V., Gilday, A., Kozhevnikova, E., et al. (2012). A cytosolic acyltransferase contributes to triacylglycerol synthesis in sucrose-rescued *Arabidopsis* seed oil catabolism mutants. *Plant Physiol.* 160, 215–225. doi: 10.1104/pp.112.201541
- Hu, Y., Li, Y., Weng, J., Liu, H., Yu, G., Liu, Y., et al. (2021). Coordinated regulation of starch synthesis in maize endosperm by microRNAs and DNA methylation. *Plant J.* 105, 108–123. doi: 10.1111/tpj.15043
- Hu, Y. F., Li, Y. P., Zhang, J., Liu, H., Tian, M., and Huang, Y. (2012). Binding of ABI4 to a CACCG motif mediates the ABA-induced expression of the ZmSSI gene in maize (*Zea mays* L.) endosperm. *J. Exp. Bot.* 63, 5979–5989. doi: 10.1093/jxb/ers246
- Jain, M., Chevala, V. V., and Garg, R. (2014). Genome-wide discovery and differential regulation of conserved and novel microRNAs in chickpea via deep sequencing. *J. Exp. Bot.* 65, 5945–5958. doi: 10.1093/jxb/eru333
- Jin, J., Tian, F., Yang, D. C., Meng, Y. Q., Kong, L., Luo, J., et al. (2017). PlantTFDB 4.0: toward a central hub for transcription factors and regulatory interactions in plants. *Nucleic Acids Res.* 45, 1040–1045. doi: 10.1093/nar/gkw982
- Jones, S. I., and Vodkin, L. O. (2013). Using RNA-Seq to profile soybean seed development from fertilization to maturity. *PLoS ONE* 8:e59270. doi: 10.1371/journal.pone.0059270
- Kang, G., Liu, G., Peng, X., Wei, L., Wang, C., Zhu, Y., et al. (2013). Increasing the starch content and grain weight of common wheat by overexpression of the cytosolic AGPase large subunit gene. *Plant Physiol. Biochem.* 73, 93–98. doi: 10.1016/j.plaphy.2013.09.003
- Klaus, D., Ohlrogge, J. B., Neuhaus, H. E., and Dörmann, P. (2004). Increased fatty acid production in potato by engineering of acetyl-CoA carboxylase. *Planta* 219, 389–396. doi: 10.1007/s00425-004-1236-3
- Konishi, T., and Sasaki, Y. (1994). Compartmentalization of two forms of acetyl-CoA carboxylase in plants and the origin of their tolerance toward herbicides. *Proc. Natl. Acad. Sci. U.S.A.* 91, 3598–3601. doi: 10.1073/pnas.91.9.3598
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Lakhssassi, N., Zhou, Z., Liu, S., Colantonio, V., AbuGhazaleh, A., and Meksem, K. (2017). Characterization of the *FAD2* gene family in soybean reveals the limitations of gel-based TILLING in genes with high copy number. *Front. Plant Sci.* 8:324. doi: 10.3389/fpls.2017.00324
- 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
- Lee, S. J., Cho, D. I., Kang, J. Y., and Kim, S. Y. (2009). An ARIA-interacting AP2 domain protein is a novel component of ABA signaling. *Mol. Cells* 27, 409–416. doi: 10.1007/s10059-009-0058-3
- Li, N., Zhang, S., Zhao, Y., Li, B., and Zhang, J. (2011). Over-expression of AGPase genes enhances seed weight and starch content in transgenic maize. *Planta* 233, 241–250. doi: 10.1007/s00425-010-1296-5
- Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Arabidopsis Book* 11:e0161. doi: 10.1199/tab.0161

- Liu, F., Xia, Y., Wu, L., Fu, D., Hayward, A., Luo, J., et al. (2015). Enhanced seed oil content by overexpressing genes related to triacylglyceride synthesis. *Gene* 557, 163–171. doi: 10.1016/j.gene.2014.12.029
- Lu, K. J., Streb, S., Meier, F., Pfister, B., and Zeeman, S. C. (2015). Molecular genetic analysis of glucan branching enzymes from plants and bacteria in *Arabidopsis* reveals marked differences in their functions and capacity to mediate starch granule formation. *Plant Physiol.* 169, 1638–1655. doi: 10.1104/pp.15.00792
- Ma, W., Kong, Q., Grix, M., Mantyla, J. J., Yang, Y., Benning, C., et al. (2015). Deletion of a C-terminal intrinsically disordered region of WRINKLED1 affects its stability and enhances oil accumulation in *Arabidopsis*. *Plant J.* 83, 864–874. doi: 10.1111/tpj.12933
- Nakamura, Y. (2002). Towards a better understanding of the metabolic system for amylopectin biosynthesis in plants: rice endosperm as a model tissue. *Plant Cell Physiol.* 43, 718–725. doi: 10.1093/pcp/pcf091
- Nayar, S., Sharma, R., Tyagi, A. K., and Kapoor, S. (2013). Functional delineation of rice *MADS29* reveals its role in embryo and endosperm development by affecting hormone homeostasis. *J. Exp. Bot.* 64, 4239–4253. doi: 10.1093/jxb/ert231
- Nicholas, K. B., Nicholas, H. B. J., and Deerfield, D. W. (1997). Genedoc: analysis and visualization of genetic variation. *Embnew News* 4:14.
- Ohdan, T., Francisco, P. B. Jr., Sawada, T., Hirose, T., Terao, T., Satoh, H., et al. (2005). Expression profiling of genes involved in starch synthesis in sink and source organs of rice. *J. Exp. Bot.* 56, 3229–3244. doi: 10.1093/jxb/eri292
- Okamuro, J. K., Caster, B., Villarroel, R., Van Montagu, M., and Jofuku, K. D. (1997). The AP2 domain of APETALA2 defines a large new family of DNA binding proteins in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* 94, 7076–7081. doi: 10.1073/pnas.94.13.7076
- Qi, X., Li, S., Zhu, Y., Zhao, Q., Zhu, D., and Yu, J. (2017). ZmDof3, a maize endosperm-specific Dof protein gene, regulates starch accumulation and aleurone development in maize endosperm. *Plant Mol. Biol.* 93, 7–20. doi: 10.1007/s11103-016-0543-y
- Rathi, D., Gayen, D., Gayali, S., Chakraborty, S., and Chakraborty, N. (2016). Legume proteomics: progress, prospects, and challenges. *Proteomics* 16, 310–327. doi: 10.1002/pmic.201500257
- Ritchie, S. W., Hanway, J. J., Thompson, H. E., and Benson, G. O. (1982). *How a soybean plant develops*. Ames, IA: Iowa State University of Science and Technology Cooperative Extension Service.
- Roesler, K., Shintani, D., Savage, L., Boddupalli, S., and Ohlrogge, J. (1997). Targeting of the *Arabidopsis* homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. *Plant Physiol.* 113, 75–81. doi: 10.1104/pp.113.1.75
- Rosli, R., Chan, P. L., Chan, K. L., Amiruddin, N., Low, E. L., Singh, R., et al. (2018). *In silico* characterization and expression profiling of the diacylglycerol acyltransferase gene family (*DGAT1*, *DGAT2*, *DGAT3* and *WS/DGAT*) from oil palm, *Elaeis guineensis*. *Plant Sci.* 275, 84–96. doi: 10.1016/j.plantsci.2018.07.011
- Routaboul, J. M., Benning, C., Bechtold, N., Caboche, M., and Lepiniec, L. (1999). The TAG1 locus of *Arabidopsis* encodes for a diacylglycerol acyltransferase. *Plant Physiol. Biochem.* 37, 831–840. doi: 10.1016/S0981-9428(99)00115-1
- Saha, S., Enugutti, B., Rajakumari, S., and Rajasekharan, R. (2006). Cytosolic triacylglycerol biosynthetic pathway in oilseeds. Molecular cloning and expression of peanut cytosolic diacylglycerol acyltransferase. *Plant Physiol.* 141, 1533–1543. doi: 10.1104/pp.106.082198
- Scheidig, A., Fröhlich, A., Schulze, S., Lloyd, J. R., and Kossmann, J. (2002). Downregulation of a chloroplast-targeted beta-amylase leads to a starch-excess phenotype in leaves. *Plant J.* 30, 581–591. doi: 10.1046/j.1365-3113X.2002.01317.x
- Schwarte, S., Wegner, F., Havenstein, K., Groth, D., Steup, M., and Tiedemann, R. (2015). Sequence variation, differential expression, and divergent evolution in starch-related genes among accessions of *Arabidopsis thaliana*. *Plant Mol. Biol.* 87, 489–519. doi: 10.1007/s11103-015-0293-2
- Shen, B., Allen, W. B., Zheng, P., Li, C., Glassman, K., Ranch, J., et al. (2010). Expression of *ZmLEC1* and *ZmWRI1* increases seed oil production in maize. *Plant Physiol.* 153, 980–987. doi: 10.1104/pp.110.157537
- Shimada, T. L., and Hara-Nishimura, I. (2010). Oil-body-membrane proteins and their physiological functions in plants. *Biol. Pharm. Bull.* 33, 360–363. doi: 10.1248/bpb.33.360
- Singer, S. D., Chen, G., Mietkiewska, E., Tomasi, P., Jayawardhane, K., Dyer, J. M., et al. (2016). *Arabidopsis* GPAT9 contributes to synthesis of intracellular glycerolipids but not surface lipids. *J. Exp. Bot.* 67, 4627–4638. doi: 10.1093/jxb/erw242
- Sun, C., Palmqvist, S., Olsson, H., Borén, M., Ahlandsberg, S., and Jansson, C. (2003). A novel WRKY transcription factor, SUSIBA2, participates in sugar signaling in barley by binding to the sugar-responsive elements of the iso1 promoter. *Plant Cell* 15, 2076–2092. doi: 10.1105/tpc.014597
- Szydlowski, N., Ragel, P., Raynaud, S., Lucas, M. M., Roldán, I., Montero, M., et al. (2009). Starch granule initiation in *Arabidopsis* requires the presence of either class IV or class III starch synthases. *Plant Cell* 21, 2443–2457. doi: 10.1105/tpc.109.066522
- Tan, H., Yang, X., Zhang, F., Zheng, X., Qu, C., Mu, J., et al. (2011). Enhanced seed oil production in canola by conditional expression of *Brassica napus* LEAFY COTYLEDON1 and LEC1-LIKE in developing seeds. *Plant Physiol.* 156, 1577–1588. doi: 10.1104/pp.111.175000
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917
- Tang, T., Du, C., Song, H., Aziz, U., Wang, L., Zhao, C., et al. (2019). Genome-wide analysis reveals the evolution and structural features of WRINKLED1 in plants. *Mol. Genet. Genomics* 294, 329–341. doi: 10.1007/s00438-018-1512-8
- Tang, X. J., Peng, C., Zhang, J., Cai, Y., You, X. M., Kong, F., et al. (2016). ADP-glucose pyrophosphorylase large subunit 2 is essential for storage substance accumulation and subunit interactions in rice endosperm. *Plant Sci.* 249, 70–83. doi: 10.1016/j.plantsci.2016.05.010
- Thelen, J. J., and Ohlrogge, J. B. (2002). Both antisense and sense expression of biotin carboxyl carrier protein isoform 2 inactivates the plastid acetyl-coenzyme A carboxylase in *Arabidopsis thaliana*. *Plant J.* 32, 419–431. doi: 10.1046/j.1365-3113X.2002.01435.x
- Tian, Z., Qian, Q., Liu, Q., Yan, M., Liu, X., Yan, C., et al. (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc. Natl. Acad. Sci. U.S.A.* 106, 21760–21765. doi: 10.1073/pnas.0912396106
- Troncoso-Ponce, M. A., Kilaru, A., Cao, X., Durrett, T. P., Fan, J., Jensen, J. K., et al. (2011). Comparative deep transcriptional profiling of four developing oilseeds. *Plant J.* 68, 1014–1027. doi: 10.1111/j.1365-3113X.2011.04751.x
- Tuncel, A., Kawaguchi, J., Ihara, Y., Matsusaka, H., Nishi, A., Nakamura, T., et al. (2014). The rice endosperm ADP-glucose pyrophosphorylase large subunit is essential for optimal catalysis and allosteric regulation of the heterotetrameric enzyme. *Plant Cell Physiol.* 55, 1169–1183. doi: 10.1093/pcp/pcu057
- Wang, H., Guo, J., Lambert, K. N., and Lin, Y. (2007). Developmental control of *Arabidopsis* seed oil biosynthesis. *Planta* 226, 773–783. doi: 10.1007/s00425-007-0524-0
- Wang, J., Sun, P., Li, Y., Liu, Y., Yu, J., Ma, X., et al. (2017). Hierarchically aligning 10 legume genomes establishes a family-Level genomics platform. *Plant Physiol.* 174, 284–300. doi: 10.1104/pp.16.01981
- Wang, J. C., Xu, H., Zhu, Y., Liu, Q. Q., and Cai, X. L. (2013). OsZIP58, a basic leucine zipper transcription factor, regulates starch biosynthesis in rice endosperm. *J. Exp. Bot.* 64, 3453–3466. doi: 10.1093/jxb/ert187
- Wang, L., Czedik-Eysenberg, A., Mertz, R. A., Si, Y., Tohge, T., Nunes-Nesi, A., et al. (2014). Comparative analyses of C4 and C3 photosynthesis in developing leaves of maize and rice. *Nat. Biotechnol.* 32, 1158–1165. doi: 10.1038/nbt.3019
- Weselake, R. J., Taylor, D. C., Rahman, M. H., Shah, S., Laroche, A., McVetty, P. B. E., et al. (2009). Increasing the flow of carbon into seed oil. *Biotechnol. Adv.* 27, 866–878. doi: 10.1016/j.biotechadv.2009.07.001
- Wu, G. Z., and Xue, H. W. (2010). *Arabidopsis* β -ketoacyl-[acyl carrier protein] synthase i is crucial for fatty acid synthesis and plays a role in chloroplast division and embryo development. *Plant Cell* 22, 3726–3744. doi: 10.1105/tpc.110.075564
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39, 316–322. doi: 10.1093/nar/gkr483
- Yang, K., Tian, Z., Chen, C., Luo, L., Zhao, B., Wang, Z., et al. (2015). Genome sequencing of adzuki bean (*Vigna angularis*) provides insight into high starch

- and low fat accumulation and domestication. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13213–13218. doi: 10.1073/pnas.1420949112
- Yang, T., Li, Y., Liu, Y., He, L., Liu, A., Wen, J., et al. (2021). The 3-ketoacyl-CoA synthase WFL is involved in lateral organ development and cuticular wax synthesis in *Medicago truncatula*. *Plant Mol. Biol.* 105, 193–204. doi: 10.1007/s11103-020-01080-1
- Yang, T., Xu, R., Chen, J., and Liu, A. (2016). β -Ketoacyl-acyl carrier protein synthase I (KASI) plays crucial roles in the plant growth and fatty acids synthesis in tobacco. *Int. J. Mol. Sci.* 17:1287. doi: 10.3390/ijms17081287
- Yao, Y., You, Q., Duan, G., Ren, J., Chu, S., Zhao, J., et al. (2020). Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. *BMC Plant Biol.* 20:51. doi: 10.1186/s12870-019-2199-7
- Zhang, J., Chen, J., Yi, Q., Hu, Y., Liu, H., Liu, Y., et al. (2014). Novel role of ZmaNAC36 in co-expression of starch synthetic genes in maize endosperm. *Plant Mol. Biol.* 84, 359–369. doi: 10.1007/s11103-013-0153-x
- Zhang, Z., Dunwell, J. M., and Zhang, Y. M. (2018). An integrated omics analysis reveals molecular mechanisms that are associated with differences in seed oil content between *Glycine max* and *Brassica napus*. *BMC Plant Biol.* 18:328. doi: 10.1186/s12870-018-1542-8
- Zhao, L., Katavic, V., Li, F., Haughn, G. W., and Kunst, L. (2010). Insertional mutant analysis reveals that long-chain acyl-CoA synthetase 1 (LACS1), but not LACS8, functionally overlaps with LACS9 in *Arabidopsis* seed oil biosynthesis. *Plant J.* 64, 1048–1058. doi: 10.1111/j.1365-313X.2010.04396.x
- Zhu, H., Choi, H. K., Cook, D. R., and Shoemaker, R. C. (2005). Bridging model and crop legumes through comparative genomics. *Plant Physiol.* 137, 1189–1196. doi: 10.1104/pp.104.058891

Conflict of Interest: 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 © 2021 Cheng, Pan, Liu, Zhang 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.



Optimizing RNAi-Target by *Nicotiana benthamiana*-Soybean Mosaic Virus System Drives Broad Resistance to Soybean Mosaic Virus in Soybean

Hua Jiang, Kai Li and Junyi Gai*

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

OPEN ACCESS

Edited by:

Madan K. Bhattacharyya,
Iowa State University, United States

Reviewed by:

Angelika Voronova,
Latvian State Forest Research
Institute Silava (LSFR), Latvia
Ahmad A. Omar,
University of Florida, United States

*Correspondence:

Junyi Gai
sri@njau.edu.cn

Specialty section:

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

Received: 15 July 2021

Accepted: 22 October 2021

Published: 22 November 2021

Citation:

Jiang H, Li K and Gai J (2021)
Optimizing RNAi-Target by *Nicotiana benthamiana*-Soybean Mosaic Virus System Drives Broad Resistance to Soybean Mosaic Virus in Soybean.
Front. Plant Sci. 12:739971.
doi: 10.3389/fpls.2021.739971

Soybean mosaic virus (SMV) is a prevalent pathogen of soybean (*Glycine max*). Pyramiding multiple SMV-resistance genes into one individual is tedious and difficult, and even if successful, the obtained multiple resistance might be broken by pathogen mutation, while targeting viral genome via host-induced gene silencing (HIGS) has potential to explore broad-spectrum resistance (BSR) to SMV. We identified five conserved target fragments (CTFs) from S1 to S5 using multiple sequence alignment of 30 SMV genome sequences and assembled the corresponding target-inverted-repeat constructs (TIRs) from S1-TIR to S5-TIR. Since the inefficiency of soybean genetic transformation hinders the function verification of batch TIRs in SMV-resistance, the *Nicotiana benthamiana*-chimeric-SMV and *N. benthamiana*-pSMV-GUS pathosystems combined with *Agrobacterium*-mediated transient expression assays were invented and used to test the efficacy of these TIRs. From that, S1-TIR assembled from 462 bp CTF-S1 with 92% conservation rate performed its best on inhibiting SMV multiplication. Accordingly, S1-TIR was transformed into SMV-susceptible soybean NN1138-2, the resistant-healthy transgenic T₁-plants were then picked out via detached-leaf inoculation assay with the stock-plants continued for progeny reproduction (T₁ dual-utilization). All the four T₃ transgenic progenies showed immunity to all the inoculated 11 SMV strains under individual or mixed inoculation, achieving a strong BSR. Thus, optimizing target for HIGS via transient *N. benthamiana*-chimeric-SMV and *N. benthamiana*-pSMV-GUS assays is crucial to drive robust resistance to SMV in soybean and the transgenic S1-TIR-lines will be a potential breeding source for SMV control in field.

Keywords: broad-spectrum resistance (BSR), host-induced gene silencing (HIGS), conserved target fragments (CTFs), soybean mosaic virus (SMV), transient assay in *N. benthamiana*, transgenic soybeans

INTRODUCTION

Soybean mosaic virus (SMV) is a prevalent viral pathogen of soybean [*Glycine max* (L.) Merr.], especially in China (Hill and Whitham, 2014; Li and Zhi, 2016); it usually leads to soybean yield reductions that varied with cultivar, virus strain, location, and infection time (Hajimorad et al., 2018). Soybean mosaic virus belongs to *Potyvirus*; its genome is a single-stranded, positive-sense (+) RNA, with a length of ~ 9.6 Kb. It is prone to high mutation rates, generating genetic diversity and different SMV strains (Seo et al., 2009b). Based on the symptoms produced on a set of different soybean genotypes, SMV isolates can be classified into different strain groups. For instance, seven SMV strains were reported in the United States based on the responses of a set of eight differential soybean genotypes to the virus (Cho and Goodman, 1979). Similarly, five SMV strains in Japan and eleven SMV strains in Korea have been reported (Takahashi et al., 1980; Seo et al., 2009a). In China, more than 4,500 country-wide SMV isolates were collected and grouped into twenty-two strains based on the responses of ten differential hosts to the virus at the National Center for Soybean Improvement (Li et al., 2010; Wang et al., 2014).

For different SMV strains, a number of SMV resistance loci were identified in host soybean-resistant sources. For example, three loci, *Rsv1*, *Rsv3*, and *Rsv4* in the United States and sixteen loci, *Rsc3*~*Rsc8*, *Rsc10*~*Rsc15*, *Rsc17*, *Rsc18*, *Rsc20*, and *Rsc21* in China have been identified from different resistant sources (Li and Zhi, 2016; Karthikeyan et al., 2017). However, pyramiding resistance genes conferring all SMV strains in a region or country into one individual using conventional methods is difficult and tedious, and it is always accompanied with undesirable traits from multiple parental sources (Malay et al., 2016). Besides, the breakdown of *R* gene-mediated resistance occurred frequently due to high mutation rates of SMV genome, interactions between the virus and its host, and the strong directional selection pressure created by the wide use of limited resistant resources (Choi et al., 2005; Gagarinova et al., 2008). In improving the situation and obtaining broad-spectrum resistance (BSR) to SMV, we have attempted to solve the SMV-disease problem based on combining host-induced gene silencing (HIGS) strategy with genetic engineering technology.

Host-induced gene silencing or host-delivered RNAi is a natural antiviral mechanism, which is initiated when double-stranded RNAs or hairpin RNAs (hpRNAs) are processed by dsRNA-specific dicer-like enzymes into 21- to 24-nucleotide (nt) short/small interfering RNAs which are then incorporated into the RNA-induced silencing complex to guide the degradation or translational repression of RNA targets in a sequence-specific manner (Yang and Li, 2018). Indeed, the hpRNAs are formed by the expression of the target-inverted-repeat (TIR) sequence ("target" is a piece of sequence of the infectious agent served as the stem in the structure of hpRNAs) in the corresponding host plant in most cases. For example, the transgenic soybean plants containing CP-TIR sequences of *Soybean dwarf virus* were developed and shown to be

resistant to *Soybean dwarf virus* (Tougou et al., 2006). The transgenic soybean plants containing CP- or HC-Pro-TIR sequences of SMV showed resistance to SMV, but only a few plants exhibited SMV resistance in T₁ generation (Kim et al., 2013, 2016; Gao et al., 2015a) while most transgenic lines transformed with SMV-CP-TIR sequences were susceptible to SMV inoculation. Although CP- or HC-Pro-targeting transgenes may be effective to a limited range of SMV strains, HIGS has shown potentials in exploring BSR to SMV. A key step for the success of HIGS strategy may lie in the identification of suitable target fragments common in numerous viral genomes since a conserved common target may cope with SMV strains with the same target sequence. However, due to low efficiency, time-consumption, and labor-inefficiency of soybean genetic transformation, to find a quick and efficient system for optimizing the best one from a mass of potential targets of HIGS before soybean genetic transformation is a key for the development of BSR to SMV.

Nicotiana benthamiana is an ideal model plant for the investigation of plant-pathogen interaction mainly due to a large number of diverse plant viruses that can successfully infect it, and *Agrobacterium*-mediated transient gene expression works exceptionally well in *N. benthamiana* plants (Norkunas et al., 2018). It has been successfully used for the identification of new viral suppressors of RNA silencing and functional analysis of unidentified genes, artificial microRNAs, or RNAi constructs to different plant viruses (Thomas et al., 2003; Peart et al., 2005; Goodin et al., 2008; Roumi et al., 2012; Medinahernández et al., 2013). In general, *N. benthamiana* plants could not be infected by SMV except several reports (Almeida et al., 2002; Gao et al., 2015b; Jiang et al., 2017). Thereinto, our previous study found that *N. benthamiana* plants can be systemically infected by an SMV of *Watermelon mosaic virus* (WMV) chimeric isolate 4278-1 (SWCI-4278-1), which has certain potential to verify the effectiveness of TIR constructs derived from the conserved target fragment (CTF) (Medinahernández et al., 2013). In addition, a recombinant SMV vector, pSMV-GUS was also considered for testing the infectivity of *N. benthamiana* plants in this study. Thus, we expect to use the two *N. benthamiana*-SMV pathosystems combined with *Agrobacterium*-mediated transient expression assay to quickly and efficiently select the possible HIGS targets, which are to be further verified for their function using a limited scale of transformed soybeans.

Accordingly, to explore the utilization of HIGS for BSR to SMV in soybeans, this study was focused on the following factors: (i) identifying CTFs throughout SMV genomes based on multiple sequence alignment and assembling TIR constructs, (ii) establishing and demonstrating a transient assay for identifying optimal TIR constructs on inhibiting SMV-multiplication using *N. benthamiana*-SWCI-4278-1 and *N. benthamiana*-pSMV-GUS pathosystems, and (iii) establishing a transgenic demonstrating system with dual-utilization of T1 in identifying SMV-resistant positive transgenic individuals and in obtaining corresponding progenies for a large-scale testing of the resistance spectrum. Based on the study, the obtained BSR materials might be used in real breeding programs.

MATERIALS AND METHODS

Plant Materials and Soybean Mosaic Virus Strains

The broad-spectrum susceptible host soybean cultivar *NN1138-2* was used throughout the study along with the transient assay host, *N. benthamiana*. The *NN1138-2* and *N. benthamiana* plants, grown at 25°C with a 16 h photoperiod in a greenhouse, were used for *Agrobacterium tumefaciens*-mediated transformation, agroinfiltration, and SMV inoculation. Altogether 11 SMV strains (SC1, SC2, SC3, SC4, SC6, SC10, SC13, SC16, SC17, SC18, and SC19) provided by the National Center for Soybean Improvement (Nanjing Agricultural University, Nanjing, China) were used in testing for resistance spectrum.

Identification of Conserved Target Fragments of Soybean Mosaic Virus Genome and Construction of Target-Inverted-Repeat Constructs

Conserved sequences are identical or have similar sequences of nucleic acids or amino acids. Sequence conservation is positively related to sequence identity. Accordingly, multiple sequence alignment was conducted using ClustalW algorithm for the complete genome sequences of 30 SMV isolates downloaded from the NCBI database¹ (Supplementary Table 1). Five CTFs (S1–S5) were selected according to sequence conservation in the sequence alignment result and the length of the CTFs approximately range from 300 to 500 bp. The sequence conservation of these fragments was analyzed using DNASP software with conserved DNA regions module in the analysis option, i.e., sequence conservation was calculated by the ratio of the number of variant sites to the net number of analyzed sites (Rozas et al., 2017). Then the sequences of the five fragments were cloned separately into pHellsgate12 (Helliwell and Waterhouse, 2003) to obtain S1-TIR, S2-TIR, S3-TIR, S4-TIR, and S5-TIR constructs, respectively; at the same time, the S1 fragment was also cloned into pB7GWIWG2(II) (Karimi et al., 2002) to obtain pB7GWIWG2(II)-S1 using gateway method. The vector pB7GWIWG2(II) contains a bar gene as the selectable marker gene (conferring resistance to herbicide glufosinate). The primers required for the construction of these vectors are listed in Supplementary Table 2. All cloned sequences were confirmed by resequencing.

Efficacy Validation of Target-Inverted-Repeat Constructs Using Transient *Nicotiana benthamiana* –SWCI-4278-1 and *Nicotiana benthamiana* –pSMV-GUS Pathosystems

From the *A. tumefaciens*-infiltrating leaves of *N. benthamiana*, eight leaves were infiltrated by each construct in one replicate, half is used for mock inoculation, and half is used for virus

inoculation, and three replicates were conducted. The constructs from S1-TIR to S5-TIR were transformed into *A. tumefaciens* strain, GV3101 through electroporation, and the infiltration method was performed according to the protocols described by Wydro et al. (2006) with some modification. Briefly, the transformed *A. tumefaciens* GV3101 was cultured in 4 ml of Luria-Bertan (LB) medium (50 µg/ml spectinomycin) at 200 rpm for approximately 34 h at 28°C. The cell cultures were centrifuged at 4,500 rpm for 5 min and then the tight pellets were gently re-suspended in 10 mM MgCl₂ (CAS:7786-30-3, Macklin Biochemical, Shanghai, China). The suspension was diluted to OD₆₀₀ = 0.5 with a solution containing a final concentration of 10 mM MgCl₂, 10 mM of 2-(N-morpholino) ethanesulfonic acid (MES) (CAS:145224-94-8, Sigma, United States) pH 5.7, and 150 µM of acetosyringone (CAS:2478-38-8, Biosharp, Hefei, China), and then infiltrated into 4-week-old leaves of *N. benthamiana* with a 1 mL needleless syringe.

Two virus inoculums were used for the inoculation of *A. tumefaciens* infiltrated leaves, where one is an SMV-WMV chimeric isolate SWCI-4278-1 and the other one is a recombinant SMV vector pSMV-GUS, i.e., a full-length of 9,994 bp SMV cDNA clone with GUS inserted between the P1 and HC-Pro cistrons were cloned into pGreen vector backbone, which was provided by the Laboratory of Plant Virology, Nanjing Agricultural University. For the inoculation of SWCI-4278-1 (Jiang et al., 2017), the SWCI-infected leaves of *N. benthamiana* were homogenized in 0.01 M phosphate buffer (PH 7.4) in a grinder and the tissue homogenates were filtered with two layers of cheesecloth. Then, the *A. tumefaciens*-infiltrated leaves of *N. benthamiana* were mechanically inoculated with a filtered sap. For the inoculation of infectious clone pSMV-GUS, the coinfiltration of infectious clone pSMV-GUS and TIR constructs was performed, i.e., an equal volume of both *Agrobacterium* cultures OD₆₀₀ = 1.0 were mixed before infiltration. The inoculated or infiltrated plants were grown in insect-free chambers with 16 h light/8-h dark cycles at 25°C.

For virus detection of the SWCI-4278-1 and pSMV-GUS inoculated leaves, the real-time fluorescence quantitative PCR (qRT-PCR), ELISA and β-glucuronidase (GUS) histochemical staining and GUS activity were performed. For qRT-PCR, total RNA was extracted from *N. benthamiana* leaves using RNA isolation kit (Tiangen, Beijing, China) at 3 and 5dpi, respectively, and the quality and concentration of RNA were checked using agarose gel electrophoresis and a spectrophotometer (Nanodrop 2000, Thermo Fisher Scientific, MA, United States). The first-strand complementary DNA (cDNA) was synthesized using PrimeScriptTM RT reagent Kit (TAKARA, Dalian, China) and qRT-PCR was performed according to previously described protocols (Jiang et al., 2019). JHA125 and JHA17 primers were used for amplifying the SMV isolate 4278-1 and pSMV-GUS, respectively; *EF1a* and *PP2A* were used as internal controls in *N. benthamiana* plants (Liu et al., 2012), and the relative accumulation of viral content was calculated by normalizing the geometric mean of the two reference genes. Each experiment was of three biological repeats, where each repeat had three technical replicates. The relative quantification of viral RNA was analyzed by $2^{-\Delta\Delta CT}$ methods based on the amplification

¹<http://www.ncbi.nlm.nih.gov/>

efficiencies (Livak and Schmittgen, 2001; Pfaffl, 2001). The above primer sequences are listed in **Supplementary Table 3**.

For ELISA assay at 3 and 5dpi, the SMV antiserum (polyclonal rabbit antibodies) of ELISA was purchased from ACD Inc. (cat #V094-R1, Beijing, China) and the instructions of the manufacturer were followed for the next steps.

For GUS-histochemical staining and activity at 5dpi, histochemical staining of *N. benthamiana* leaves was performed using 5-bromo-4-chloro-3-indolyl glucuronide (X-gluc) (CAS:114162-64-0, War biological Technology, Nanjing, China) as described by Jefferson (1987). For fluorometric assays, the plant tissues were homogenized in an extraction buffer [50 mM phosphate buffer (pH 7.0), 10 mM of beta-mercaptoethanol (CAS:60-24-2, Solarbio Science & Technology, Beijing, China), 10 mM EDTA (CAS:6381-92-6, Solarbio Science & Technology, Beijing, China), 0.1% (w/v) of sodium dodecyl sulfate (CAS: 151-21-3, Solarbio Science & Technology, Beijing, China), 0.1% (V/V) Triton X-100 (CAS:9002-93-1, Solarbio Science & Technology, Beijing, China)], and centrifuged at 12,000 g for 15 min at 4°C. Enzyme reactions were performed in 200 µl of extraction buffer containing 1 mM of 4-methylumbelliferyl-β-d-glucuronide (CAS:6815-91-4, Shanghai, China) at 37°C, and were stopped by the addition of 20 µl reaction buffer and 180 µl of 0.2 M Na₂CO₃ (CAS:497-19-8, Solarbio Science & Technology, Beijing, China). Protein concentrations were quantified using the method described by Bradford (1976). The 4-methylumbelliferone fluorescence was measured using a SpectraMax M5 (Molecular Devices, Sunnyvale, CA, United States). GUS activities were expressed in nanomoles of 4-methylumbelliferone released per minute per milligram total protein.

Northern Blot Analysis, Reverse Transcription-Polymerase Chain Reaction, and Western Blot Analysis

Total RNA was extracted from *N. benthamiana* leaves infiltrated with *A. tumefaciens*, non-transgenic and transgenic soybean leaves, respectively, using the Trizol reagent (Cat no. 15596-026, Invitrogen, Carlsbad, CA, United States) following the instructions of the manufacturer. The RNA specimens were electrophoresed on 15% polyacrylamide/7 M urea gels and transferred to Hybond N⁺ membranes (Amersham) using a semidry blotter (Bio-rad, Hercules, CA, United States). The membranes were UV cross-linked in a HL-2000 HybriLinkerTM (AnalytikJena: UVP) at 120,000 µJ/cm² energy for 4 min and separately hybridized with five different digoxin-UTP-labeled RNA probes. The mouse anti-digoxin monoclonal antibody (1:10000, Jackson ImmunoResearch) and IRDye 800CW-conjugated goat (polyclonal) anti-mouse IgG (1:15000; H + L; LI-COR Biosciences) secondary antibodies were used to evaluate siRNA production. The membranes were visualized using an LICOR Odyssey scanner with excitation at 700 and 800 nm. For Reverse Transcription-Polymerase Chain Reaction (RT-PCR) assay, the 320 bp SMV CP and 1,812 bp GUS bands were amplified from pSMV-GUS-infiltrated *N. benthamiana* leaves. The Western blot analyses were performed according to

previously described protocols and the size of CP protein is ~ 30 kDa (Jiang et al., 2019).

Off-Target Prediction of S1 Sequences, Soybean Transformation, and Identification of Transgenic Soybean Plants

The off-target searching of S1 sequences in soybean *Williams* 82 transcript database was conducted by siRNA-Finder software (Lück et al., 2019). The pB7GWIWG2(II)-S1 was transformed into *A. tumefaciens* strain EHA105 by electroporation; the *A. tumefaciens*-mediated soybean genetic transformation was carried out according to the method described by Li et al. (2017). Genomic PCR, LibertyLink[®] strip detection (QuickStixTM Kit purchased from EnviroLogix Inc., cat #AS 013 LS, Portland, ME, United States), and Northern blot were used to confirm gene insertion into soybean genome and gene expression. In addition, 492 bp S1, 557 bp *Tubulin*, and 399 bp *bar* bands were amplified from transgenic soybean plants. The primer sequences are listed in **Supplementary Table 3**. The genomic PCR primers were listed in **Supplementary Table 3**. LibertyLink[®] strip detection was performed according to the instructions of the manufacturer. The siRNAs produced from the supposed transgenic soybean plants were detected using the Northern Blot method in the present study.

Virus Inoculation of Detached Leaves *in vitro* and Transgenic Soybean Plants

When the unifoliate leaves of non-transgenic and transgenic soybean plants were fully expanded, one unifoliate leaf of each plant was split into two along the main vein and half of the unifoliate leaf was used for SMV strain SC18 inoculation. The SC18-infected leaves of soybean were homogenized in 0.01 M phosphate buffer (PH 7.4) in a grinder and the tissue homogenates were filtered with two layers of cheesecloth. Then, half of the unifoliate leaf was mechanically inoculated by a paintbrush using the filtered sap with a small amount of 600-mesh carborundum powder (CAS:409-21-2, Sinopharm Chemical Reagent, Shanghai, China). The SMV-inoculated leaves were placed in sterile petri dishes containing a wet filter paper and further covered with a piece of wet sterile cotton in the area near the petiole. The plates were then wrapped in a parafilm and placed in a growth-chamber at 25°C under 16 h photoperiod culture. After 7 days, the leaves were used for SMV detection through ELISA. For inoculation of non-transgenic and transgenic soybean plants, the 11 SMV strains were separately increased and maintained on a highly susceptible soybean *NN1138-2* and the SMV-infected soybean leaves were homogenized in 0.01 M phosphate buffer (PH 7.4) in different grinders and the tissue homogenates were filtered with two layers of cheesecloth, respectively. Then, the fully developed unifoliate leaves of soybean plants were mechanically inoculated with the above-mentioned filtered sap, respectively. For the mix inoculation, the individual saps of the same volume were mixed for inoculation. The number of plants used for each inoculation

treatment is 20. The inoculated plants were grown in insect-free chambers with 16 h light/8 h dark cycles at 25°C, which were further used for resistance evaluation. ELISA detection assay was also performed at 28 dpi as per the instructions of the manufacturer.

Statistical Analyses

The ELISA and qRT-PCR data were analyzed for their means, SD, and single-factor analyses of variance using the SPSS software (version 18.0, SPSS Inc., Chicago, IL, United States). Duncan's new multiple range tests were performed to determine any significant difference among various treatments using the significance level of $\alpha = 0.05$.

The S1 nucleotide sequences were obtained from SMV (taxid: 12222) nucleotide collection of NCBI database, which were used for retrieving the highly similar sequences in the NCBI database by BLASTN suite with default parameter values, and the SMV genome sequences with significant alignments (the sequence identity is greater than 80%) were selected.

RESULTS

Identification of Conserved Target Fragments in Soybean Mosaic Virus Genome

Focusing on the assumption of selecting relatively conserved sequences of SMV genome as the target of HIGS may contribute to the success of BSR to SMV, the potential target sequences in the SMV genome were analyzed through multiple sequence alignment of the complete genome sequences of 30 SMV isolates (Supplementary Table 1). The five CTFs (named S1 to S5) were identified and distributed at different positions on the SMV genome (Supplementary Figure 1A). The five CTFs, S1 to S5, varied in length from 322 to 497 bp with relatively low G/C and high A/U content (Table 1). The sequence conservation rates were 92%, 84%, 83%, 83%, and 86%, respectively, and varied with numbers of the conservative and variable sites among the fragments (Table 1). Meanwhile, the sequence conservation rate among the 30 SMV genomes was only ~28%, far lower than that of the selected CTFs. Furthermore, S1-TIR to S5-TIR assembled from CTF S1 to S5 was separately transiently expressed in *N. benthamiana* leaves, and the siRNAs produced from these constructs were detected by Northern blotting with respective DIG-labeled RNA probes. Supplementary Figure 1B shows the production of the siRNAs derived from five different CTFs and indicates that all the five TIR constructs can be expressed normally in *N. benthamiana* plants.

Differential Inhibition of SWCI-4278-1 Accumulation by Five Target-Inverted-Repeat-Constructs in *Nicotiana benthamiana* Leaves

In a previous study, it was found that the *N. benthamiana* plants could be systemically infected by an SMV-WMV chimeric isolate, i.e., SWCI-4278-1 (Jiang et al., 2017). It is not known, whether

N. benthamiana-SWCI-4278-1 pathosystem can be used to optimize TIR constructs for inhibiting SMV multiplication. Thus, the S1-TIR, S2-TIR, S3-TIR, S4-TIR, and S5-TIR constructs with vector control were transiently expressed in *N. benthamiana* leaves by agroinfiltration, and further mechanically inoculated with SWCI-4278-1. The response and virus content were surveyed for the inoculated leaves at 3- and 5-days post-inoculation (dpi), respectively. No symptoms were developed on the inoculated leaves. The qRT-PCR results of the isolate SWCI-4278-1 nucleic acid is shown in Table 2. The amount of viral RNA was ~12.8%–14.8%, ~24.2%–32.5%, ~21.6%–27.5%, ~17.4%–20.4% and ~14.8%–16.3% in *N. benthamiana* leaves for S1-TIR, S2-TIR, S3-TIR, S4-TIR, and S5-TIR constructs, respectively, compared to the vector control (Table 2, left side). The significance tests showed that the S1-TIR was similar to the S5-TIR and superior to other three TIRs in the inhibition of viral nucleic acid accumulation at 5dpi (Table 2, left side). Consistent with the qRT-PCR analysis, the ELISA results revealed that the viral coat protein accumulated at lower levels (an average of ~14.7%–29.4% at 3dpi and ~18.0%–34.7% at 5dpi) in S1-TIR, S2-TIR, S3-TIR, S4-TIR, and S5-TIR-expressed leaves compared to the vector control expressed leaves (Table 3). It was especially true for S1-TIR and S5-TIR according to Duncan's new multiple range tests ($\alpha = 0.05$). Collectively, these data indicated that the *N. benthamiana*-SWCI-4278-1 pathosystem combined with *Agrobacterium*-mediated transient expression assay worked exceptionally well in the screening of high-efficient TIR constructs. Thus SWCI-4278-1 infection was significantly inhibited in the inoculated leaves of *N. benthamiana* due to the expressed five different TIR constructs and the inhibition effect of S1-TIR and S5-TIR were superior over the other three TIR constructs on virus accumulation.

Viral Multiplication Was Inhibited by Five Target-Inverted-Repeat Constructs via Transient *Nicotiana benthamiana*-pSMV-GUS Pathosystem

For testing the infectivity of a recombinant vector pSMV-GUS in *N. benthamiana* plants, the phenotype, the viral nucleic acid, and coat protein of the pSMV-GUS infiltrated-*N. benthamiana* plants were assessed at 7, 14, and 21 dpi, respectively. No symptoms were observed on the inoculated leaves and in the non-inoculated systemic leaves. However, the RT-PCR analysis confirmed the existence of the viral nucleic acids in *N. benthamiana* plants infiltrated with pSMV-GUS using viral and GUS-specific primers, respectively, which indicates that the pSMV-GUS successfully replicated in the inoculated leaves and further spread to the non-inoculated leaves of *N. benthamiana* plants (Figure 1A). Meanwhile, the viral coat protein was detected in the systemic leaves of *N. benthamiana* plants by Western blotting, which confirmed the upward movement of pSMV-GUS in *N. benthamiana* plants (Figure 1B). Thus, *N. benthamiana* plants could be systemically infected by pSMV-GUS. Importantly, the friendly recombinant SMV infectious clone, pSMV-GUS can be used as an inoculum for testing the effect of the TIR constructs to SMV infection.

TABLE 1 | Characteristics of the identified five conserved target fragments in Soybean mosaic virus (SMV) genome based on multiple sequence alignment.

Fragment name	Location [†]	Fragment length (bp)	G + C content (%) [‡]	A + U content (%) [§]	Number of variable sites	Number of conserved sites	Sequence conservation (%)
S1	2595–3056	462	39	61	38	426	92
S2	5601–6013	413	45	55	65	348	84
S3	6930–7426	497	39	61	83	414	83
S4	8261–8582	322	42	58	54	268	83
S5	8919–9274	356	42	58	50	306	86

[†] The different nucleotide positions of the conserved target fragments (CTFs) on SMV genome.

[‡] The proportion of guanine and cytosine nucleotides in total nucleotides of the different CTFs, respectively.

[§] The proportion of adenine and uracil nucleotides in total nucleotides of the different CTFs, respectively.

TABLE 2 | Relative quantification of viral RNA in *N. benthamiana* leaves for different TIR constructs.

Sample name	SWCI-4278-1 as inoculum			Mock RQ	pSMV-GUS as inoculum		
	RQ (mean ± SD)		Percentage [†] 3dpi/5dpi (%)		RQ (mean ± SD)		Percentage [†] 3dpi/5dpi (%)
	3dpi	5dpi			3dpi	5dpi	
EV	1085.65 ± 114.57a	2165.11 ± 172.53a	–	1.00	748.76 ± 21.28a	1504 ± 36.47a	–
S1-TIR	161.15 ± 9.90c	277.10 ± 25.59d	14.8/12.8	1.06	30.22 ± 3.34b	71.45 ± 1.66d	4.0/4.8
S2-TIR	352.35 ± 35.44b	523.39 ± 69.69b	32.5/24.2	0.94	45.12 ± 2.95b	121.62 ± 4.89b	6.0/8.1
S3-TIR	299.25 ± 26.31b	467.37 ± 39.16b	27.5/21.6	1.16	39.93 ± 1.43b	103.72 ± 4.50bc	5.3/6.9
S4-TIR	188.47 ± 20.56c	440.84 ± 27.58bc	17.4/20.4	1.00	34.35 ± 3.05b	83.23 ± 9.13cd	4.6/5.5
S5-TIR	177.43 ± 19.48c	319.99 ± 32.46cd	16.3/14.8	1.01	33.53 ± 1.74b	76.55 ± 6.53cd	4.5/5.1

The mean and SD were calculated from three replications.

Values in the same column followed by different letters are significantly different at $p = 0.05$ level. RQ: relative quantification; SD: standard deviation.

[†] Percentage was determined by the RQ means of each TIR sample divided by the RQ means of EV sample at 3dpi and 5dpi, respectively.

SWCI: SMV/WMV-chimeric isolate; EV: empty vector; TIR: target-inverted-repeat; dpi: days post-inoculation; -: no data; S1-TIR to S5-TIR represent the 5 respective TIR constructs; Mock: 0.01 M phosphate buffer inoculation as control.

TABLE 3 | The virus content in *N. benthamiana* leaves from DAS-ELISA analysis for different TIR constructs.

Sample name	ELISA (mean ± SD)			
	Mock	SWCI-4278-1		Percentage [†] 3dpi/5dpi (%)
		3dpi	5dpi	
EV	0.11 ± 0.01	1.02 ± 0.13a	1.67 ± 0.14a	–
S1-TIR	0.10 ± 0.01	0.15 ± 0.02c	0.30 ± 0.02d	14.7/18.0
S2-TIR	0.11 ± 0.01	0.30 ± 0.04b	0.58 ± 0.04b	29.4/34.7
S3-TIR	0.12 ± 0.01	0.17 ± 0.02c	0.47 ± 0.04bc	16.7/28.1
S4-TIR	0.10 ± 0.01	0.16 ± 0.02c	0.44 ± 0.04c	15.7/26.3
S5-TIR	0.12 ± 0.01	0.15 ± 0.02c	0.31 ± 0.02d	14.7/18.6

The mean and standard deviation were calculated from three replications.

Values in the same column followed by different letters are significantly different at $p = 0.05$ level. SD: standard deviation.

[†] Percentage was determined by the ELISA mean of each TIR sample divided by the ELISA mean of EV sample at 3dpi and 5dpi, respectively.

SWCI: SMV/WMV-chimeric isolate; EV: empty vector; TIR: target-inverted-repeat; dpi: days post-inoculation; -: no data; S1-TIR to S5-TIR represent the TIR constructs; Mock: 0.01 M phosphate buffer inoculation as control.

Next, whether pSMV-GUS infection can be restricted in *N. benthamiana* leaves by the five TIR constructs was investigated. *A. tumefaciens* containing individual different TIR constructs were mixed with the *A. tumefaciens*-contained pSMV-GUS in an equal proportion, respectively. The mixture was further infiltrated into *N. benthamiana* leaves through the injection method. As shown in **Table 2**, the relative content of pSMV-GUS nucleic acids was measured at 3dpi and 5dpi,

respectively, using qRT-PCR. The amount of virus accumulation was ~4.0%–4.8%, ~6.0%–8.1%, ~5.3%–6.9%, ~4.6%–5.5%, and ~4.5%–5.1% in *N. benthamiana* leaves for S1-TIR, S2-TIR, S3-TIR, S4-TIR, and S5-TIR constructs, respectively, compared to vector control-infiltrated leaves (**Table 2**, right side). The five TIR constructs performed similarly on the inhibition level of pSMV-GUS multiplication at 3 dpi, while the S1-TIR was superior to S2-TIR and S3-TIR and similar to S4-TIR and S5-TIR at

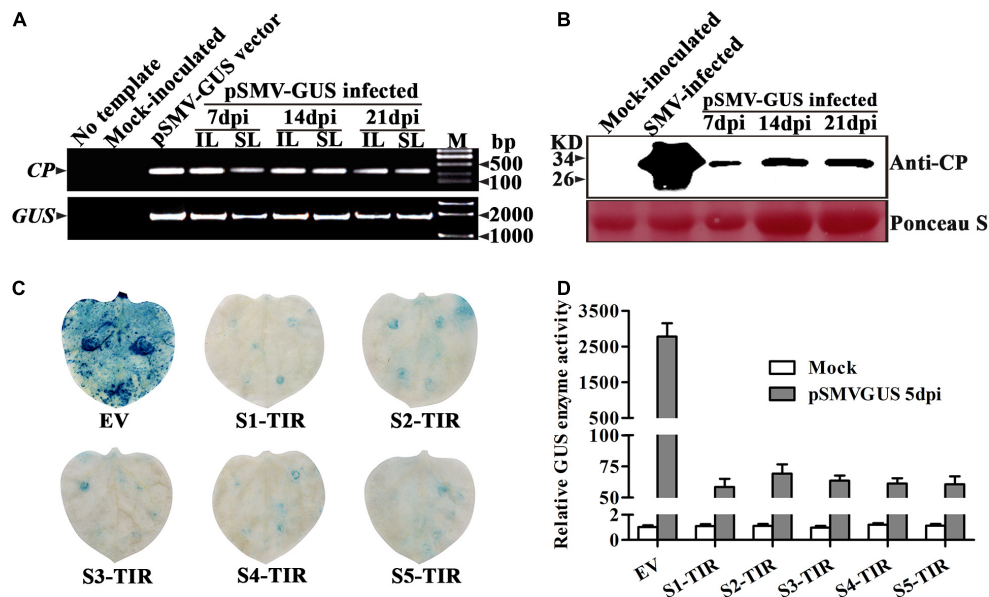


FIGURE 1 | Systemic infection of recombinant SMV cDNA construct pSMV-GUS in *Nicotiana Benthamiana* plants and detection of pSMV-GUS accumulation in *N. benthamiana* leaves expressing S1-TIR to S5-TIR. **(A)** Detection of viral nucleic acid in inoculated and systemic leaves of *N. benthamiana* plants by RT-PCR. IL: inoculated leaves, SL: systemic leaves. **(B)** Western blot confirmed the presence of CP protein in the systemic leaves of *N. benthamiana* plants at 7, 14, and 21 dpi, respectively. **(C)** GUS histochemical staining of the inoculated leaves of *N. benthamiana* at 5 dpi. **(D)** GUS enzyme activity analysis of inoculated leaves infiltrated by different TIR constructs at 5 dpi. EV: empty vector, TIR: target-inverted-repeat, S1-TIR to S5-TIR represent the respective TIR constructs. Mock: 0.01 M phosphate buffer inoculation as control. The error bars indicate SD calculated from three biological replications.

5dpi (Table 2, right side). Moreover, the GUS histochemical staining for all different inoculated leaves was observed at 5dpi (Figure 1C). As expected, the leaves coinfiltrated with vector control and pSMV-GUS displayed significantly higher GUS accumulation than the leaves coinfiltrated with each TIR construct and pSMV-GUS (Figure 1C), which indicated less virus accumulation in the five TIR constructs-expressed leaves. Meanwhile, the GUS enzyme activity test was in accordance with the observed histochemical analysis (Figure 1D). According to the above results, the efficacy of five TIR constructs on inhibiting virus multiplication was successfully validated based on transient *N. benthamiana*-pSMV-GUS pathosystem. From the above, the results of two transient *N. benthamiana*-SWCI-4278-1/pSMV-GUS assays both indicated that the CTF-S1 and CTF-S5 are more effective targets for *N. benthamiana*-induced gene silencing than the others, while between the two CTFs, CTF-S1 should be more potential because of its higher sequence conservation (Figure 1 and Tables 1–3). Thus, the S1-TIR will be transformed into SMV-susceptible soybean, and the resistance of transgenic soybeans to SMV strains will be further evaluated.

Generation of S1-TIR Transgenic Soybeans and Detached Leaf-Assay for Identifying Soybean Mosaic Virus Resistant T₁-Plants

Before S1-TIR genetic transformation of soybean, the off-target searching of S1 sequences in soybean *Williams 82* transcript database was conducted by siRNA-Finder software and no targets

were found. The soybean genetic transformation assay obtained twelve independent T₀ plants, which were transferred to the greenhouse, where one plant died and eleven survived. The S1-TIR insertion, bar protein existence, and siRNA production were analyzed for the randomly selected six transgenic T₀ plants, and the results confirmed that the positive S1-TIR transgenic soybean plants were obtained (Supplementary Figure 2). All eleven transgenic T₀ plants were self-fertilized to generate T₁ generation.

In order to screen out positive SMV-resistant individual plants from T₁ generation and to avoid the potential effect of virus inoculation on transgenic plant growth and reproduction, one plant for two purposes (or dual functions) were adopted, i.e., SMV-resistance testing based on detached leaf-assay and the positive stock plants for generation extension and seed reproduction. The resistance-testing results of the randomly selected twenty-four T₁ plants are graphically displayed in Figure 2. The virus content of half of a detached-unifoliate leaf inoculated with SMV SC18 sap was detected using ELISA at 7 dpi (Figure 2B, down side), and the other half of the detached-unifoliate leaf was used for S1-TIR construct insertion detection (Figure 2B, upper side). The results showed that the virus content is different among each detached-unifoliate leaf (Figure 2B). All T₁ individual plants were tested based on the detached leaf-assay, and the positive rate and resistance rate of each transgenic line were counted (Table 4). In the T₁ transgenic soybean plants, the positive rate was between 61.8 and 77.1%, and the resistance rate was between 28.3 and 60.5% (Table 4). The positive rate varied slightly while the resistance

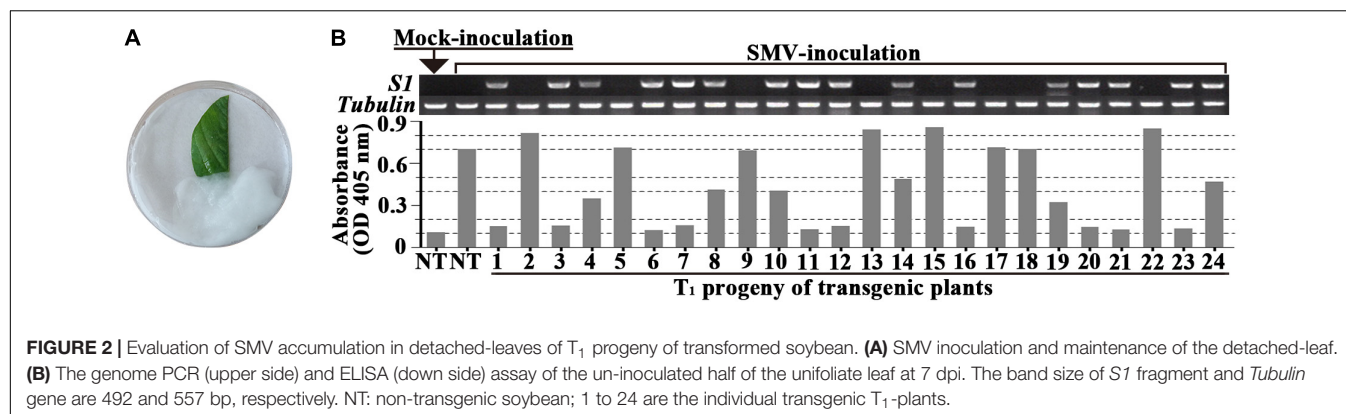


TABLE 4 | Transgene-positive and resistant ratios in T_1 generation from *in vitro* leaf-assay.

T_0 code	Seed no.	No. of individual plants obtained	Transgene-positive plant		Resistant and susceptible plant [§]		R(%) [¶]
			No. [†]	(%) [‡]	(+)	(-)	
L1	100	89	60	67.4	43	17	28.3
L3	86	85	57	67.1	29	28	49.1
L4	17	17	11	64.7	6	5	45.5
L5	163	149	99	66.4	56	43	43.4
L6	255	223	172	77.1	68	104	60.5
L7	101	97	61	62.9	25	36	59.0
L8	65	60	39	65.0	20	19	48.7
L9	103	102	63	61.8	32	31	49.2
L10	28	23	17	73.9	7	10	58.8
L11	72	67	42	62.7	27	15	35.7
L12	108	92	57	62.0	27	30	52.6
Total	1098	1004	678	66.5	340	338	48.3

[†] The number of transgene-positive plants, identification of transgene-positive plants was based on genome DNA PCR using gene-specific primers.

[‡] The percentage of obtained T_1 plants.

[§] Assessment of the resistant or susceptible plants was based on the ratio of the ELISA mean of each transgene-positive individual under SMV SC18 inoculation to the ELISA mean of negative control non-inoculated at 7dpi using *in vitro* leaf-assay. (+): positive for SMV (ratio ≥ 2); (-): negative for SMV (ratio < 2).

[¶] The percentage of resistant plants.

rate varied greatly. The results indicated that the differences in the proportion of resistant individual plants among different transgenic lines.

Moreover, the higher ratio was found in four transgenic lines (L6, L7, L10, and L12) (Table 4). Taken together, a total of 338 T_1 SMV-resistant positive plants were obtained based on the detached leaf-assay. In brief, the developed approach is useful for preliminary screening SMV-strain resistance in individual plants.

The S1-TIR Transgene Confers Robust Resistance to Multiple Soybean Mosaic Virus Strains in T_3 Lines

In order to evaluate the performance of the whole plant of the transgenic progeny to multiple SMV strains, four T_3 transgenic lines (L6, L7, L10, and L12) were selected and challenged with eleven SMV strains (SC1, SC2, SC3, SC4, SC6, SC10, SC13, SC16, SC17, SC18, and SC19), as control buffers and wild-type plants (susceptible soybean variety NN1138-2) were used. Figure 3 shows the responses of wild-type and transgenic plants

to virus challenge. The typical symptoms of SMV include mosaic, mottling, and curling of soybean leaves, which may be different among strains. As shown in Figure 3A, the wild-type plants showed significant SMV symptoms on systemic leaves, while L6 transgenic plants showed a complete absence of symptoms for the eleven different SMV strains infection which was indistinguishable from the mock-inoculated controls at 28dpi (Supplementary Table 4). Similar results were also obtained for transgenic soybean plants L7, L10, and L12 after virus inoculation, confirming that SMV resistance can be seen in all four independent lines (Supplementary Table 4). Furthermore, the extracts of systemic leaves from individual plants were assayed by ELISA. As shown in Figure 3B, the SMV coat protein was undetectable in all transgenic plants inoculated with eleven SMV strains, confirming virus resistance.

Moreover, since the naturally mixed infection of SMV strains on soybean plants may occur in the field, T_3 transgenic soybean plants were also tested with a mixed inoculum of eleven SMV strains (Figure 3B). No viral symptoms were detected on these four transgenic lines upon infectivity assay which

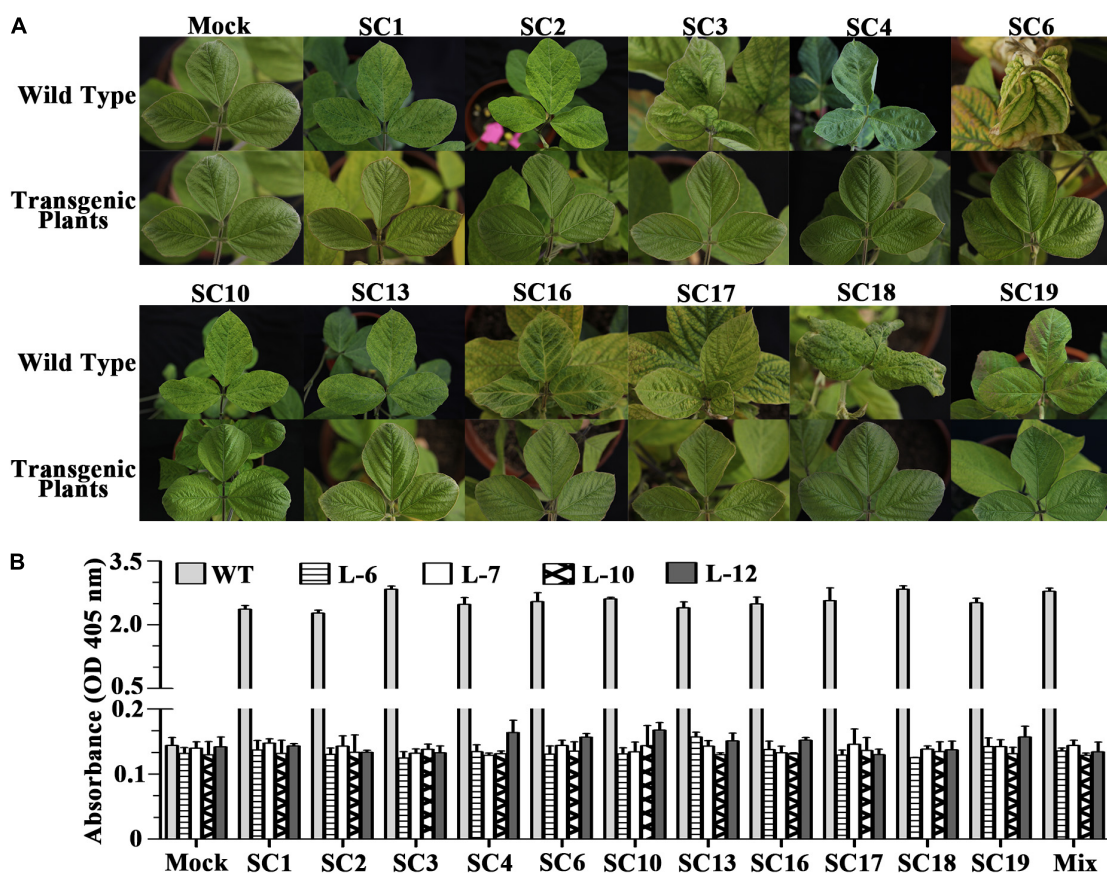


FIGURE 3 | Responses of the non-transgenic and T_3 transgenic soybean plants to SMV inoculation. **(A)** The photographs of the systemic leaves of SMV-inoculated non-transgenic and T_3 transgenic plants were taken at 28 dpi. Wild Type = receptor soybean variety NN1138-2; Transgenic plants are those of Line L-6; Mock: non-inoculated plants as controls. SC1, SC2, SC3, SC4, SC6, SC10, SC13, SC16, SC17, SC18, and SC19: different SMV strains. **(B)** ELISA detection of SMV coat protein in the systemic leaves of SMV-inoculated non-transgenic (WT = NN1138-2) and four transgenic lines (L-6, L-7, L-10, and L-12) at 28 dpi. Mix: a mixed inoculum of the above eleven SMV strains. The error bars indicate standard deviations with $N = 20$.

were confirmed further by ELISA analysis (Figure 3B). Thus, the S1-TIR transgene conferred resistance to mixed infection of the eleven SMV strains in transgenic soybeans. Besides, the similarity analysis of *S1* among a wide range of SMV sequences showed that the identity ranged from 82.3 to 99.8% within SMV isolates from China, from 92.0 to 95.7% in Korea, from 92.0 to 97.8% with those from the United States, from 92.6 to 94.6% with those from Canada, from 93.5 to 97.6% with those from Iran, and from 82.8 to 95% with those from United Kingdom (Supplementary Table 5). It is especially important that the identity ranges from 91.8 to 99.8% in those SMV isolates collected from *Glycine* hosts, except several isolates collected from some atypical SMV hosts (Supplementary Table 5).

In brief, the CTF-*S1* obtained from SMV genome was demonstrated to be an efficient target of HIGS and did drive the development of BSR to SMV in soybean, so as to have a great potential in the breeding of novel soybean varieties with comprehensive resistance to various SMV strains or isolates and can benefit the field control of the virus.

DISCUSSION

Identification of Conserved Target Fragment in Soybean Mosaic Virus Genome Is the Key for the Success of Host-Induced Gene Silencing-Mediated Broad-Spectrum Resistance

In this study, five CTFs for HIGS in SMV genome were identified based on the 30 SMV genome sequences, covering different source countries, with sequence conservation rate of 92%, 84%, 83%, 83%, and 86% for genome sequences from *S1* to *S5*, respectively. The results of transient *N. benthamiana*-SMV assays implied that the higher is the sequence conservation, the better is the working of HIGS. The higher sequence conservation of the CTF means it is more common among SMV genome sequences, and its TIR may recognize more extensive SMV strains and mediated BSR to SMV infection. Indeed, the CTF-*S1* not only showed very high nucleotide acid sequence identity (91.8 to 99.8%) with the complete genome sequences of about 50 strains/isolates isolated from cultivated and wild soybeans, but

it also showed from 82.3 to 96.5% nucleotide acid sequence identity with the complete genome sequences of several SMV isolates collected from some atypical SMV hosts. It implies that S1-TIR might be of more potential in inducing BSR to more SMV strains or isolates with more effectiveness in the host and atypical hosts. That means the CTF-S1 is a highly conservative segment among SMV genomes and S1-TIR might be potential in developing BSR to SMV.

On the other hand, the sequence conservation of a CTF may be related to its length and number of sequence pieces. The previous results showed that TIR constructs with different lengths of hairpin viral target sequence have been successfully used in various plant species (Das and Sherif, 2020). Both short (~50 bp) and long (up to 2.5 Kbp) length of the stem in the hairpin structure were used (Lee et al., 2017). Integrating multiple small fragments with sequence conservation 100% as the CTF of HIGS might be a novel option for interference with more SMV strains, while it needs to be verified by experiments. In addition, the range of G/C content of our five CTFs was ~39%–45%, meeting the criterion of 30%–52% G/C content (Fakhr et al., 2016). The highly G-rich sequences in the selection of siRNAs sequences should be avoided because they tend to form G-quartet structures. It explains why our five CTFs were all effective in using HIGS to induce the resistance to SMV, especially the BSR to SMV of CTF-S1.

***Nicotiana benthamiana* –SWCI-4278-1/ pSMV-GUS Pathosystems Form the Key for Rapid and Efficient Identification of Best Target-Inverted-Repeat Constructs**

Several approaches for transiently identifying gene function in soybean and *N. benthamiana* have been reported in recent studies (Willman, 2015; Jiang et al., 2017, 2019; Wu and Hanzawa, 2018). For instance, a method of infiltrating soybean leaves with *Agrobacterium* was developed for transiently expressing DNA of interest, while the inadequacy of replicated experiments and the number of infectious tissues might have limited their practical implications (Willman, 2015). Wu and Hanzawa (2018) established a method for the isolation of soybean protoplasts and its application to transient gene expression studies, but this approach needs optimizing the conditions since many factors that affect protoplast yields and transformation rates were involved. Additionally, two pathosystems for studying soybean–virus interactions were developed, i.e., ARISHR-SMV (*Agrobacterium rhizogenes*-induced soybean hairy roots versus SMV) and *N. benthamiana*–SWCI-4278-1 pathosystems (Jiang et al., 2017, 2019). The former was demonstrated effective for at least nine SMV strains, while the latter takes less time for the functional analysis of unidentified genes since the transient expression in *N. benthamiana* leaves was rapid and reliable (Norkunas et al., 2018), and the leaves inoculated with the virus using mechanical friction were simple and convenient. In the present study, the virus accumulation was significantly inhibited in the leaves of *N. benthamiana* transiently expressed by five TIR constructs, and the inhibition effect of S1-TIR and S5-TIR were superior to others while

S2-TIR showed the weakest inhibition ability. The effect comparison of different S1-TIR to S5-TIR constructs in virus inhibition using *N. benthamiana*–SWCI-4278-1 pathosystem confirmed that this approach is useful for identifying the best TIR construct.

Moreover, *N. benthamiana* plants can be systemically infected by a recombinant SMV infectious clone, i.e., pSMV-GUS; one possible reason is that extra 409 bp sequences of 5' terminal region in SMV genome of pSMV-GUS is highly similar to that of *Bean common mosaic virus*, which also supported that the 5'-UTR and P1 region of the *Potyvirus* play important roles in the interaction between viruses and host (Zhang et al., 2015). Thus, the novel transient *N. benthamiana*–pSMV-GUS pathosystem can be also used for identifying the efficient target of HIGS. In comparison to SWCI-4278-1 as inoculum, the leaves agro-infiltrated with five TIR constructs accumulated less virus when using pSMV-GUS as inoculum, which is probably due to the fact that the viral replication level and/or the ability of HC-Pro suppressing posttranscriptional gene silencing were different between SWCI-4278-1 and pSMV-GUS. In brief, the transient *N. benthamiana*–SWCI-4278-1 and *N. benthamiana*–pSMV-GUS assays take only several days for optimizing the best TIR construct assembled from viral CTFs and the procedure was rapid and reliable. In contrast, if the effect verification of these TIR constructs to SMV was conducted by stable soybean genetic transformation, it will take a lot of time and workload. Thus, the advantage of saving time and effort will become more obvious when testing multiple TIR constructs using *N. benthamiana*–SMV pathosystem compared to the developing and testing of transgenic soybean plants. However, the reason that *N. benthamiana* can be infected by modified SMV (SWCI-4278-1 and pSMV-GUS) needs to be further explored.

Dual-Utilization of Transgenic T₁ Plants Is the Key for Goal-Directed Obtaining of Transgenic Resistant Individuals and Their Progenies

The detached leaf-assay has been used for evaluating the plant resistance to different pathogens (Aregbesola et al., 2020; Sun et al., 2020). In this study, for avoiding the influence of the inoculated SMV damaging to plant growth and reproduction, and the potential risk of passing the inoculated SMV to the next generation transgenic plants and for saving the workload of the transgenic seed reproduction, the detached leaf-assay was used for identifying SMV-resistant positive transgenic plants with the corresponding resistant stock plants used for seed reproduction. This dual utilization method of transgenic T₁-plants is efficient and effective. This procedure may be used for rapidly screening transformed soybean materials at an early stage, which was used for filtering out the false positive plants and SMV-susceptible transgenic plants. Without this dual-utilization of T₁-plants, many more transgenic plants have to be obtained and used in identifying their resistance. Of course, this procedure may have some issues out of expectation in SMV-resistance identification of non-transgenic soybean germplasm resources, such as the movement of the virus to the system leaves may be blocked in

some soybean varieties, where even the inoculated-leaves can be infected by SMV (Wu and Cheng, 2020; Kumar and Dasgupta, 2021). In this case, the results of SMV resistance from the detached-leaves might appear false negative. In addition, In the T₁ transgenic soybean plants, the positive rate varied slightly while the resistance rate varied greatly, which might be due to the insertion site or the expression level differences of S1-TIR among individuals. Anyway, the detached leaf-assay is still a rapid method to identify SMV resistance in soybean.

The Integrated Host-Induced Gene Silencing Strategy Leads to Realizing the Broad-Spectrum Resistance to Soybean Mosaic Virus and Providing Novel Sources for Breeding Programs

Host-induced gene silencing has emerged as a powerful genetic tool in the fundamental research for the assessment of gene function and in applied research for plant protection and agronomic traits improvement, especially for viral control. In previous studies, most TIR-transgenic soybeans were susceptible to SMV inoculation; only a few plants exhibited SMV resistance in T₁ generation (Kim et al., 2013). Similarly, Gao et al. (2015a) also reported that the TIR-transgenic soybean plants with resistance to a certain SMV strain was improved using partial SMV-*HC-Pro* sequences as the target for HIGS. Different from the previous studies, the present study revealed that all the four transgenic soybean lines that expressed the S1-TIR construct optimized via transient *N. benthamiana* assays were immune to all the inoculated 11 SMV strains. It implies that S1-TIR-induced resistance is really broad, strong, and stable. In short, the obtained 11 transgenic lines are of important value to be used in breeding programs coping with the SMV damage.

Recently, the clustered regulatory interspaced short palindromic repeats and associated protein system (CRISPR/Cas) were developed for editing specific-target genome sequences. Thereinto, it has been applied to different plants for resistance to virus infection, including cassava (Mehta et al., 2019), potato (Zhan et al., 2019), *N. benthamiana* (Aman et al., 2018), and so on. Some of them were able to successfully confer resistance to the target virus and others not. Moreover, a few research revealed that the CRISPR method with the potential to speed up virus evolution should be carefully assessed as they pose significant biosafety risks (Mehta et al., 2019). In contrast, HIGS-mediated resistance to the virus may be more efficient and environmentally safer (Ali et al., 2020; Das and Sherif, 2020). It seems to be a smart approach with a greater potential for developing the SMV-BSR transgenic soybeans because the action mechanism of HIGS enables it to target multiple SMV strains simultaneously, especially for adopting the CTF as the target of HIGS.

CONCLUSION

In summary, CTF-S1 as the target of HIGS was successful in developing BSR to SMV for soybean. In the integrated

HIGS approach, high CTF conservation-rate, transient high-efficient verification of TIR via *N. benthamiana*-SWCI-4278-1 and *N. benthamiana*-pSMV-GUS pathosystems, and dual utilization of transgenic T₁-plants are the key links in realizing BSR to SMV in soybean. In addition, the obtained transgenic S1-TIR-lines are novel potential sources for BSR to SMV in breeding for SMV-resistance.

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.

AUTHOR CONTRIBUTIONS

JG and HJ conceived and designed the method and experiments. HJ and KL performed the experiments. HJ analyzed the data. JG contributed reagents, materials, and interpretation of the results. All authors read and approved the final manuscript.

FUNDING

This work was financially supported by the grants from the National Natural Science Foundation of China (31801388), the Fundamental Research Funds for the Central Universities (KJQN201909), the Key Transgenic Breeding Program of China (2014ZX08010-005B), the National Key R & D Program for Crop Breeding in China (2017YFD0101500), the MOE 111 Project (B08025), the MOE Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT_17R55), the MOA CARS-04 program, the Jiangsu Higher Education PAPD Program, and the Jiangsu JCIC-MCP.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.739971/full#supplementary-material>

Supplementary Figure 1 | The choice of SMV conserved fragments and verification of TIR constructs expression. **(A)** S1 – S5 distributed non-uniformly at different positions within SMV genome. S1: 2595 – 3056 nt, S2: 5601 – 6013 nt, S3: 6930 – 7426 nt, S4: 8261 – 8582 nt, and S5: 8919 – 9274 nt. **(B)** Expression of TIR constructs was verified in *N. benthamiana* leaves through transiently assay by Northern blot with different DIG-labeled RNA probes, respectively. EV: empty vector, S1-TIR to S5-TIR represent the TIR constructs.

Supplementary Figure 2 | Identification of the positive transgenic soybean plants. **(A)** Genome DNA PCR amplification of the 492 bp S1 and 399 bp *bar* gene in transformed soybean. -: ddH₂O as a template; NT: non-transgenic soybean; +: vector control; L1, L3, L4, L5, L6, and L9 were T₀ transgenic soybean lines. **(B)** LibertyLink® strip detection. The first strip is the control line, and the second strip is the test line. **(C)** Northern blot analysis of siRNA isolated from T₀ transgenic soybean lines (L1, L3, L4, L5, L6, and L9) detected by DIG-labeled S1 probe.

REFERENCES

- Ali, M., Javaid, A., Naqvi, S. H., Batcho, A., Kayani, W. K., Lal, A., et al. (2020). Biotic stress triggered small RNA and RNAi defense response in plants. *Mol. Biol. Rep.* 47, 5511–5522. doi: 10.1007/s10333-020-05583-4
- Almeida, A. M., Sakai, J., Souto, E. R., Kitajima, E. W., Fukuji, T. S., and Hanada, K. (2002). Mosaic in *Senna occidentalis* in Southern Brazil induced by a new strain of Soybean mosaic virus. *Fitopatol. Bras.* 27, 151–156. doi: 10.1590/S0100-41582002000200005
- Aman, R., Ali, Z., Butt, H., Mahas, A., Aljedaani, F., Khan, M. Z., et al. (2018). RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* 19:1. doi: 10.1186/s13059-017-1381-1
- Aregbesola, E., Ortega-Beltran, A., Falade, T., Jonathan, G., Hearne, S., and Bandyopadhyay, R. (2020). A detached leaf assay to rapidly screen for resistance of maize to *Bipolaris maydis*, the causal agent of southern corn leaf blight. *Eur. J. Plant Pathol.* 156, 133–145. doi: 10.1007/s10658-019-01870-4
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* 72, 248–254.
- Cho, E. K., and Goodman, R. M. (1979). Strains of Soybean mosaic virus: classification based on virulence in resistant soybean cultivars. *Phytopathology* 69, 467–490. doi: 10.1094/Phyto-69-467
- Choi, B. K., Koo, J. M., Ahn, H. J., Yum, H. J., Choi, C. W., Ryu, K. H., et al. (2005). Emergence of Rsv-resistance breaking Soybean mosaic virus isolates from Korean soybean cultivars. *Virus Res.* 112, 42–51. doi: 10.1016/j.virusres.2005.03.020
- Das, P. R., and Sherif, S. M. (2020). Application of exogenous dsRNAs-induced RNAi in agriculture: challenges and triumphs. *Front. Plant Sci.* 11:946. doi: 10.3389/fpls.2020.00946
- Fakhr, E., Zare, F., and Teimoori-Toolabi, L. (2016). Precise and efficient siRNA design: a key point in competent gene silencing. *Cancer Gene Ther.* 23, 73–82. doi: 10.1038/cgt.2016.4
- Gagarinova, A. G., Babu, M., Poysa, V., Hill, J. H., and Wang, A. M. (2008). Identification and molecular characterization of two naturally occurring Soybean mosaic virus isolates that are closely related but differ in their ability to overcome Rsv4 resistance. *Virus Res.* 138, 50–56. doi: 10.1016/j.virusres.2008.08.010
- Gao, L., Ding, X. N., Li, K., Liao, W. L., Zhong, Y. K., Ren, R., et al. (2015a). Characterization of Soybean mosaic virus resistance derived from inverted repeat-SMV-HC-Pro genes in multiple soybean cultivars. *Theor. Appl. Genet.* 128, 1489–1505. doi: 10.1007/s00122-015-2522-0
- Gao, L., Zhai, R., Zhong, Y. K., Karthikeyan, A., Ren, R., Zhang, K., et al. (2015b). Screening Isolates of Soybean mosaic virus for Infectivity in a Model Plant, *Nicotiana benthamiana*. *Plant Dis.* 99, 442–446. doi: 10.1094/PDIS-04-14-0405-RE
- Goodin, M. M., Zaitlin, D., Naidu, R. A., and Lommel, S. A. (2008). *Nicotiana benthamiana*: its history and future as a model for plant–pathogen interactions. *Mol. Plant Microbe Interact.* 21, 1015–1026. doi: 10.1094/MPMI-21-8-1015
- Hajimorad, M. R., Domier, L. L., Tolin, S. A., Whitham, S. A., and Maroof, M. A. S. (2018). Soybean mosaic virus: a successful potyvirus with a wide distribution but restricted natural host range. *Mol. Plant Pathol.* 19, 1563–1579. doi: 10.1111/mpp.12644
- Helliwell, C., and Waterhouse, P. (2003). Constructs and methods for high-throughput gene silencing in plants. *Methods* 30, 289–295. doi: 10.1016/S1046-2023(03)00036-7
- Hill, J. H., and Whitham, S. A. (2014). Control of virus diseases in soybeans. *Adv. Virus Res.* 90, 355–390. doi: 10.1016/B978-0-12-801246-8.00007-X
- Jefferson, R. A. (1987). Assaying chimeric genes in plants: the GUS gene fusion system. *Plant Mol. Biol. Rep.* 5, 387–405. doi: 10.1007/BF02667740
- Jiang, H., Li, K., Dou, D., and Gai, J. (2017). Characterization of a soybean mosaic virus variant causing different diseases in *Glycine max* and *Nicotiana benthamiana*. *Arch. Virol.* 162, 549–553. doi: 10.1007/s00705-016-3123-1
- Jiang, H., Li, K., and Gai, J. (2019). *Agrobacterium rhizogenes*-induced soybean hairy roots versus Soybean mosaic virus (ARISHR-SMV) is an efficient pathosystem for studying soybean–virus interactions. *Plant Methods* 15:56. doi: 10.1186/s13007-019-0442-8
- Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAY™ vectors for *Agrobacterium*-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3
- Karthikeyan, A., Li, K., Jiang, H., Ren, R., Li, C., Zhi, H. J., et al. (2017). Inheritance, fine-mapping, and candidate gene analyses of resistance to Soybean mosaic virus strain SC5 in soybean. *Mol. Genet. Genomics* 292, 811–822. doi: 10.1007/s00438-017-1310-8
- Kim, H. J., Kim, M.-J., Pak, J. H., Im, H. H., Lee, D. H., Kim, K.-H., et al. (2016). RNAi-mediated Soybean mosaic virus (SMV) resistance of a Korean Soybean cultivar. *Plant Biotechnol. Rep.* 10, 257–267. doi: 10.1007/s11816-016-0402-y
- Kim, H. J., Kim, M.-J., Pak, J. H., Jung, H. W., Choi, H. K., Lee, Y.-H., et al. (2013). Characterization of SMV resistance of soybean produced by genetic transformation of SMV-CP gene in RNAi. *Plant Biotechnol. Rep.* 7, 425–433. doi: 10.1007/s11816-013-0279-y
- Kumar, G., and Dasgupta, I. (2021). Variability, Functions and Interactions of Plant Virus Movement Proteins: What Do We Know So Far? *Microorganisms* 9:695. doi: 10.3390/microorganisms9040695
- Lee, G.-H., Lee, G.-S., and Park, Y.-D. (2017). New hairpin RNAi vector with *Brassica rapa* ssp. *pekinensis* intron for gene silencing in plants.
- Li, K., Yang, Q. H., Zhi, H. J., and Gai, J. Y. (2010). Identification and distribution of Soybean mosaic virus strains in southern China. *Plant Dis.* 94, 351–357. doi: 10.1094/PDIS-94-3-0351
- Li, K., and Zhi, H. J. (2016). Advances in resistance to Soybean mosaic virus disease in soybean. *Soybean Science* 35, 525–530.
- Li, S., Cong, Y., Liu, Y., Wang, T., Shuai, Q., Chen, N., et al. (2017). Optimization of *Agrobacterium*-Mediated Transformation in Soybean. *Front. Plant Sci.* 8:246. doi: 10.3389/fpls.2017.00246
- Liu, D., Shi, L., Han, C., Yu, J., Li, D., and Zhang, Y. (2012). Validation of reference genes for gene expression studies in virus-infected *Nicotiana benthamiana* using quantitative real-time PCR. *PLoS One* 7:e46451. doi: 10.1371/journal.pone.0046451
- 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
- Lück, S., Kreszies, T., Strickert, M., Schweizer, P., Kuhlmann, M., and Douchkov, D. (2019). siRNA-Finder (si-Fi) software for RNAi-target design and off-target prediction. *Front. Plant Sci.* 10:1023. doi: 10.3389/fpls.2019.01023
- Malav, A., Chandrawat, K., and Chandrawat, K. (2016). Gene pyramiding: An overview. *International Journal of Curr. Res. Biosci. Plant Biol.* 3, 22–28. doi: 10.20546/ijcrbp.2016.307.004
- Medinahernández, D., Riverabustamante, R. F., Tenllado, F., and Holguínpeña, R. J. (2013). Effects and effectiveness of two RNAi constructs for resistance to Pepper golden mosaic virus in *Nicotiana benthamiana* plants. *Viruses* 5:2931. doi: 10.3390/v5122931
- Mehta, D., Stürchler, A., Anjanappa, R. B., Zaidi, S. S.-E.-A., Hirsch-Hoffmann, M., Grussem, W., et al. (2019). Linking CRISPR-Cas9 interference in cassava to the evolution of editing-resistant geminiviruses. *Genome Biol.* 20:80. doi: 10.1186/s13059-019-1678-3
- Norkunas, K., Harding, R., Dale, J., and Dugdale, B. (2018). Improving agroinfiltration-based transient gene expression in *Nicotiana benthamiana*. *Plant Methods* 14, 1–14. doi: 10.1186/s13007-018-0343-2
- Peart, J. R., Mestre, P., Lu, R., Malcuit, I., and Baulcombe, D. C. (2005). NRG1, a CC-NB-LRR protein, together with N, a TIR-NB-LRR protein, mediates resistance against tobacco mosaic virus. *Curr. Biol.* 15, 968–973. doi: 10.1016/j.cub.2005.04.053
- Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45. doi: 10.1093/nar/29.9.e45
- Roumi, V., Afsharifar, A., Saldarelli, P., Niaz, A., Martelli, G., and Izadpanah, K. (2012). Transient expression of artificial microRNAs confers resistance to Grapevine virus A in *Nicotiana benthamiana*. *J. Plant Pathol.* 94, 643–649.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248
- Seo, J.-K., Ohshima, K., Lee, H.-G., Son, M., Choi, H.-S., Lee, S.-H., et al. (2009b). Molecular variability and genetic structure of the population of Soybean mosaic virus based on the analysis of complete genome sequences. *Virology* 393, 91–103. doi: 10.1016/j.virol.2009.07.007

- Seo, J. K., Lee, H. G., Choi, H. S., Lee, S. H., and Kim, K. H. (2009a). Infectious in vivo Transcripts from a Full-length Clone of Soybean mosaic virus Strain G5H. *Plant Pathol. J.* 25, 54–61. doi: 10.5423/PPJ.2009.25.1.054
- Sun, M., Jing, Y., Zhao, X., Teng, W., Qiu, L., Zheng, H., et al. (2020). Genome-wide association study of partial resistance to sclerotinia stem rot of cultivated soybean based on the detached leaf method. *PLoS One* 15:e0233366. doi: 10.1371/journal.pone.0233366
- Takahashi, K., Tanaka, T., Iida, W., and Tsuda, Y. (1980). Studies on virus diseases and causal viruses of soybean in Japan. *Bull. Tohoku Natl. Agric. Exp. Stn.* 62, 1–130. doi: 10.1620/tjem.130.273
- Thomas, C. L., Leh, V., Lederer, C., and Maule, A. J. (2003). Turnip crinkle virus coat protein mediates suppression of RNA silencing in *Nicotiana benthamiana*. *Virology* 306, 33–41. doi: 10.1016/S0042-6822(02)00018-1
- Tougou, M., Furutani, N., Yamagishi, N., Shizukawa, Y., Takahata, Y., and Hidaka, S. (2006). Development of resistant transgenic soybeans with inverted repeat-coat protein genes of soybean dwarf virus. *Plant Cell Rep.* 25, 1213–1218. doi: 10.1007/s00299-006-0186-6
- Wang, D. G., Li, H. W., Zhi, H. J., Tian, Z., Hu, C., Hu, G. Y., et al. (2014). Identification of strains and screening of resistance resources to Soybean mosaic virus in Anhui Province. *Chin. J. Oil Crop Sci.* 36, 374–379.
- Willman, M. (2015). *Transient Expression Efficiency of a Novel Agrobacterium Strain in Soybean Leaf Tissue*. Denman Undergraduate Research Forum (2015-03-25). Available online at: <http://hdl.handle.net/1811/67984>
- Wu, F. Q., and Hanzawa, Y. (2018). A simple method for isolation of soybean protoplasts and application to transient gene expression analyses. *J. Vis. Exp.* 131:e57258. doi: 10.3791/57258
- Wu, X., and Cheng, X. (2020). Intercellular movement of plant RNA viruses: targeting replication complexes to the plasmodesma for both accuracy and efficiency. *Traffic* 21, 725–736. doi: 10.1111/tra.12768
- Wydro, M., Kozubek, E., and Lehmann, P. (2006). Optimization of transient Agrobacterium-mediated gene expression system in leaves of *Nicotiana benthamiana*. *Acta Biochim. Pol. Engl. Ed.* 53:289. doi: 10.18388/abp.2006_3341
- Yang, Z., and Li, Y. (2018). Dissection of RNAi-based antiviral immunity in plants. *Curr. Opin. Virol.* 32, 88–99. doi: 10.1016/j.coviro.2018.08.003
- Zhan, X., Zhang, F., Zhong, Z., Chen, R., Wang, Y., Chang, L., et al. (2019). Generation of virus-resistant potato plants by RNA genome targeting. *Plant Biotechnol. J.* 17, 1814–1822. doi: 10.1111/pbi.13102
- Zhang, J., Roberts, R., and Rakotondrafara, A. M. (2015). The role of the 5' untranslated regions of Potyviridae in translation. *Virus Res.* 206, 74–81. doi: 10.1016/j.virusres.2015.02.005

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 © 2021 Jiang, Li 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.



Mutation of *GmA1TR* Genes by CRISPR/Cas9 Genome Editing Results in Enhanced Salinity Stress Tolerance in Soybean

Tianya Wang^{1†}, Hongwei Xun^{1,2†}, Wei Wang³, Xiaoyang Ding², Hainan Tian¹, Saddam Hussain¹, Qianli Dong¹, Yingying Li¹, Yuxin Cheng¹, Chen Wang¹, Rao Lin¹, Guimin Li³, Xueyan Qian², Jinsong Pang¹, Xianzhong Feng⁴, Yingshan Dong², Bao Liu¹ and Shucai Wang^{3*}

¹ Key Laboratory of Molecular Epigenetics of MOE, Northeast Normal University, Changchun, China, ² National Engineering Research Center for Soybean, Soybean Research Institute, Jilin Academy of Agricultural Sciences, Changchun, China, ³ Laboratory of Plant Molecular Genetics and Crop Gene Editing, School of Life Sciences, Linyi University, Linyi, China, ⁴ Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

OPEN ACCESS

Edited by:

Ruslan Kalendar,
University of Helsinki, Finland

Reviewed by:

Raghavendran Partha,
GRAIL Inc., United States
Zhao-Shi Xu,
Institute of Crop Sciences, Chinese
Academy of Agricultural Sciences
(CAAS), China

*Correspondence:

Shucai Wang
wangshucai@yahoo.com

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 19 September 2021

Accepted: 05 November 2021

Published: 26 November 2021

Citation:

Wang T, Xun H, Wang W, Ding X,
Tian H, Hussain S, Dong Q, Li Y,
Cheng Y, Wang C, Lin R, Li G, Qian X,
Pang J, Feng X, Dong Y, Liu B and
Wang S (2021) Mutation of *GmA1TR*
Genes by CRISPR/Cas9 Genome
Editing Results in Enhanced Salinity
Stress Tolerance in Soybean.
Front. Plant Sci. 12:779598.
doi: 10.3389/fpls.2021.779598

Breeding of stress-tolerant plants is able to improve crop yield under stress conditions, whereas CRISPR/Cas9 genome editing has been shown to be an efficient way for molecular breeding to improve agronomic traits including stress tolerance in crops. However, genes can be targeted for genome editing to enhance crop abiotic stress tolerance remained largely unidentified. We have previously identified abscisic acid (ABA)-induced transcription repressors (*A1TRs*) as a novel family of transcription factors that are involved in the regulation of ABA signaling, and we found that knockout of the entire family of *A1TR* genes in Arabidopsis enhanced drought and salinity tolerance without fitness costs. Considering that *A1TRs* are conserved in angiosperms, *A1TRs* in crops may be targeted for genome editing to improve abiotic stress tolerance. We report here that mutation of *GmA1TR* genes by CRISPR/Cas9 genome editing leads to enhanced salinity tolerance in soybean. By using quantitative RT-PCR analysis, we found that the expression levels of *GmA1TRs* were increased in response to ABA and salt treatments. Transfection assays in soybean protoplasts show that *GmA1TRs* are nucleus proteins, and have transcriptional repression activities. By using CRISPR/Cas9 to target the six *GmA1TRs* simultaneously, we successfully generated Cas9-free *gmaitr36* double and *gmaitr23456* quintuple mutants. We found that ABA sensitivity in these mutants was increased. Consistent with this, ABA responses of some ABA signaling key regulator genes in the *gmaitr* mutants were altered. In both seed germination and seedling growth assays, the *gmaitr* mutants showed enhanced salt tolerance. Most importantly, enhanced salinity tolerance in the mutant plants was also observed in the field experiments. These results suggest that mutation of *GmA1TR* genes by CRISPR/Cas9 is an efficient way to improve salinity tolerance in soybean.

Keywords: *GmA1TRs*, salinity tolerance, ABA, CRISPR/Cas9, genome editing, soybean

INTRODUCTION

As the fourth major crop and a nitrogen-fixing plant, soybean (*Glycine max*) is one of the most important protein- and oil-rich seed crops worldwide (Zhang et al., 2015; Vanllyodan et al., 2017), and it plays an important role in maintaining the cycling of nitrogen in ecosystems (Deshmukh et al., 2014). However, similar to other crops, growth and yield of soybean is largely affected by abiotic stresses including drought, salinity and extreme temperatures. As an example, drought alone can cause up to 40% yield loss of soybean globally (Wang et al., 2003; Fujita et al., 2006; Manavalan et al., 2009; Ray et al., 2013). In addition, drought and salinity are common in many different regions, and long-term drought caused by accelerated climate changes and global warming usually led to salinity. As a result, more than 50% of all arable lands on the earth may get seriously salinized by the year 2050, a dramatically increase from a currently ~20% (Wang et al., 2003). Considering that the world population is continuing increasing and an estimated increase of 70% in crop yield is needed to feed the population by the year 2050 (Ray et al., 2013; Vanllyodan et al., 2017), crop breeding to enhance abiotic stress tolerance is a critical way to improve crop yield. However, traditional breeding to improve abiotic stress tolerance may take years to decades (Manavalan et al., 2009).

Molecular breeding is able to shorten the time required for crop breeding, and the outcomes are usually more predictable compared to traditional breeding (Xu et al., 2012). The application of new developed techniques such as CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 (CRISPR-associated protein 9) genome editing in molecular breeding may further shorten the time required for crop breeding (Chen et al., 2019; Matres et al., 2021), as CRISPR/Cas9 genome editing not only enables to generate predictable mutations, but also enables to isolate transgene-free mutants from the edited transgenic plants (Ma et al., 2015; Gao et al., 2016; Lu et al., 2017; He et al., 2018; Chen et al., 2019). Since its successful application in plants (Li et al., 2013; Nekrasov et al., 2013; Shan et al., 2013), CRISPR/Cas9 genome editing has been used to improve agronomic traits in crops such as rice, tomato and wheat by editing specific target genes (Shimatani et al., 2017; He et al., 2018; Zsögon et al., 2018; Chen et al., 2019). However, identification of appropriate target genes that can be used to improve abiotic stress tolerance in crops by CRISPR/Cas9 genome editing is a challenge.

It is well known that ABA (abscisic acid) is a key stress hormone, through the PYR1/PYL/RCAR (Pyrabactin resistance 1/PYR1-like/Regulatory component of ABA receptor) receptors, the A-group PP2Cs (PROTEIN PHOSPHATASE 2C) phosphatases, the SnRKs [NON-FERMENTING 1 (SNF1)-RELATED PROTEIN KINASES] kinases, and the ABF/AREB/ABI5-type bZIP (basic region leucine zipper) transcription factors, ABA regulates the expression of ABA responsive genes and thereby plant responses to abiotic stresses such drought, salinity, cold, and heat (Rodriguez et al., 1998; Gosti et al., 1999; Fujii et al., 2007; Fujii and Zhu, 2009; Umezawa et al., 2010; Guo et al., 2011; Rushton et al., 2012; Yoshida et al., 2014; Dong et al., 2015; Tian et al., 2017). As a result, expression level changes of the ABA signaling regulator genes usually led to

changes in plant tolerance to abiotic stresses, but in most of the cases, enhanced abiotic stress tolerance was observed in plants overexpressing the regulator genes, whereas loss-of-function of the regulator genes led to reduced abiotic stress tolerance in plants (Fujita et al., 2009; Park et al., 2015; Yoshida et al., 2015; Zhao et al., 2016). Therefore, it is unlikely for the ABA signaling key regulator genes to be served as target for CRISPR/Cas9 genome editing to improve abiotic stress tolerance in crops.

We have previously identified AITRs (ABA-induced transcription repressors) as a novel family of transcription factors that function as feedback regulators of ABA signaling, and loss-of-function of *AITR* genes led to reduced ABA sensitivity in Arabidopsis (Tian et al., 2017). Consistent with the functions of AITRs in regulating ABA signaling, expression level changes of the *AITR* genes in Arabidopsis also led to changes in plant tolerance to abiotic stresses (Tian et al., 2017; Chen et al., 2021). However, different from most of the ABA signaling key regulator genes, loss-of-function of *AITR* genes resulted in enhanced tolerance to abiotic stresses including drought and salinity, whereas overexpression of *AITR5* led to reduced tolerance to salt stress in Arabidopsis (Song et al., 2016; Tian et al., 2017; Chen et al., 2021). Most importantly, knock-out-of all the six *AITR* genes in Arabidopsis led to enhanced tolerance to drought and salinity without fitness cost (Chen et al., 2021). AITRs are conserved in angiosperms, and our preliminary studies have shown that AITRs from soybean, tomato, rice and cotton shared similar features of the Arabidopsis AITRs, i.e., they are all function as transcription repressors as examined in transfected Arabidopsis protoplasts, and their expression was induced by ABA treatment (Tian et al., 2017; Wang et al., 2021). In addition, expression of a cotton *AITR* gene recovered the abiotic stress tolerance phenotypes observed in the Arabidopsis *aitr2* mutant (Wang et al., 2021), indicating that crop AITRs may have similar functions as Arabidopsis AITRs. These results suggest that AITRs may serve as CRISPR/Cas9 genome editing targets to improve abiotic stress tolerance in crops.

We report here the characterization of soybean AITRs (GmAITRs). We found that expression of GmAITRs is induced by both ABA and salt, and GmAITRs function as transcription repressors in transfected soybean protoplasts. We generated transgene-free *gmaitr* mutants by using CRISPR/Cas9 genome editing to target GmAITR genes, and found that the *gmaitr* mutants showed enhanced tolerance to salt in both laboratory and field assessments.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Williams 82 (Wm82) wild type soybean (*Glycine max*) was used for plant transformation, protoplasts isolation and as control for the experiments. The transgene-free *gmaitr36* double and *gmaitr23456* quintuple mutants were generated by using CRISPR/Cas9 gene editing in the Wm82 wild type background.

For generation assays, ABA and salt tolerance assays, and gene expression in response to ABA and salt, seeds of the Wm82 wild type and the *gmaitr* mutants were generated on the surface of two

layers of wet filter papers in Petri plates or in plastic growth bags (PhytoTC, Beijing) (Li et al., 2019), and grown in a growth room. For gene expression pattern assays and protoplast isolation, seeds were germinated in soil pots and grown in a growth room. For gene expression in response to ABA and salt, or ABA signaling key regulator gene expression, seeds were germinated and grown hydroponically in distilled water. The conditions at the growth room were set at 25°C, with 16 h light/8 h dark light cycle with light density at $\sim 600 \mu\text{mol m}^{-2} \text{s}^{-1}$, and a 60% relative humidity.

For field production analysis, seeds of the Wm82 wild type and the *gmaitr* mutants were sown and grown in three experimental fields in Jilin province, including two fields with normal soil and one saline-alkali soil field, i.e., normal soil field 1 (E124°48', N43°30'), normal soil field 2 (E125°05', N43°44'), and the saline-alkali soil field (E122°45', N45°20'), in the year 2020. The saline-alkali soil field is a typical saline-alkali land with pH 8.1–9.8, and total soluble salt 0.1–0.7%.

Sequence Alignment, Conserved Motif Analysis, and Three-Dimensional Protein Structure Prediction of *GmAITRs*

The full-length amino acid sequences of the six *GmAITRs* identified previously (Tian et al., 2017), were subjected to amino acid sequence alignment by using BioEdit with default settings, to motif analysis by using MEME¹ with default settings (Bailey et al., 2009). The *GmAITR* sequences in the Wm82 wild type and the *gmaitr* mutants were used for three-dimensional protein structures prediction by using AlphaFold v2.0² with default settings (Jumper et al., 2021). The protein structural alignment and root mean square deviation (RMSD) values were analyzed by PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.). The protein structure of *GmAITR2* was drawn by BIOVIA Discovery Studio Visualizer 2020.³

Phylogenetic Analysis

The full-length amino acid sequences of *GmAITRs*, or *GmAITRs* and AITRs from Arabidopsis, Medicago, and rice were used for alignment on MAFFT⁴ (Katoh and Standley, 2013). Phylogenetic tree was generated based on the sequence alignment result, by using MEGA7 (Kumar et al., 2016). The cross-species analysis of AITRs was performed by using the Neighbor-Joining method based on the Poisson correction substitution model. All ambiguous positions were removed for each sequence pair. The sequences used in phylogenetic analysis have been listed by Tian et al. (2017).

Abscisic Acid and NaCl Treatment

To examine the expression of *GmAITRs* in response to ABA and NaCl, healthy and uniform-sized seeds of the Wm82 wild type were selected and grown hydroponically in distilled water for 14 days. The seedlings were then transferred to 100 μM ABA,

200 mM NaCl or distilled water as a control, and treated for 6 h. Roots and leaves were dissected from the seedlings immediately after the treatments, frozen in liquid nitrogen and stored in -80°C for RNA extraction.

To examine ABA response of the ABA signaling key regulator genes, seeds of the Wm82 wild type and the *gmaitr* mutants grown in the plastic bags with distilled water for 14 days, then transferred to 100 μM ABA and distilled water as a control, and treated for 6 h. After the treatments, roots were collected and frozen in liquid nitrogen and stored in -80°C for RNA extraction.

RNA Isolation, cDNA Synthesis and qRT-PCR

For ABA response of *GmAITR* genes and ABA signaling key regulator genes, the above mentioned samples collected were used for RNA isolation. For tissue expression analysis, roots, stems and leaves were collected from 28-day-old soil pot-grown Wm82 wild type plants when the trifoliate leaf fully opened, frozen in liquid nitrogen and stored in -80°C for RNA extraction.

Total RNA was isolated from the samples collected by using an OminiPlant RNA kit (CWBIO) according to the manufacturer's instructions. During the isolation, RNA was treated with RNase-Free DNase (CWBIO) to avoid the contamination of DNA. After the DNase treatment, 1 or 2 μg total RNA was used to synthesize cDNA by oligo(dT)20-primed reverse transcription using the EasyScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech). The synthesized cDNA was used as the template for gene expression analysis. For qRT-PCR, each sample was amplified in three parallel reactions as technical replicates, and the *GmEF-1 α* (*Glyma.17G186600*) was amplified as a reference gene. The primers used for genes *GmPYL9*, *GmPYL10*, *GmPYL12*, and *GmPP2C1* have been described previously (Bai et al., 2013), and the primers used for expression analysis of other genes are listed in **Supplementary Table 1**.

Constructs

The reporter construct *LexA-Gal4:GUS*, and the effector constructs *GD*, *GD-GmAITRs*, *GFP-GmAITRs*, and *LD-VP* have been described previously (Tiwari et al., 2004; Wang et al., 2005; Tian et al., 2017).

To generate CRISPR/Cas9 constructs for *GmAITRs* gene editing, the potential target sequences within the exons of *GmAITRs* were selected by using targetDesign on CRISPR-GE.⁵ Target specificity was then evaluated by using offTarget on CRISPR-GE. A total of six target sequences were selected. Due to the high CDS sequence similarity (>85%) between *GmAITR* gene pairs, i.e., *GmAITR1* and *GmAITR4*, *GmAITR2* and *GmAITR5*, and *GmAITR3* and *GmAITR6*, each of the six target sequences was able to target one pair of genes. The six targets were divided into two groups with each group contains three target sequences that can target all the six genes. The target sequences were inserted into the *pYL-CRISPR/Cas9P_{ubi}-B* vector to generate CRISPR/Cas9 genome editing constructs using the method described previously (Ma et al., 2015). The target sequences in construct one are

¹<http://meme-suite.org>

²<https://www.alphafold.ebi.ac.uk/>

³<https://www.3ds.com/>

⁴<https://mafft.cbrc.jp/alignment/server/>

⁵<http://skl.scau.edu.cn/home/>

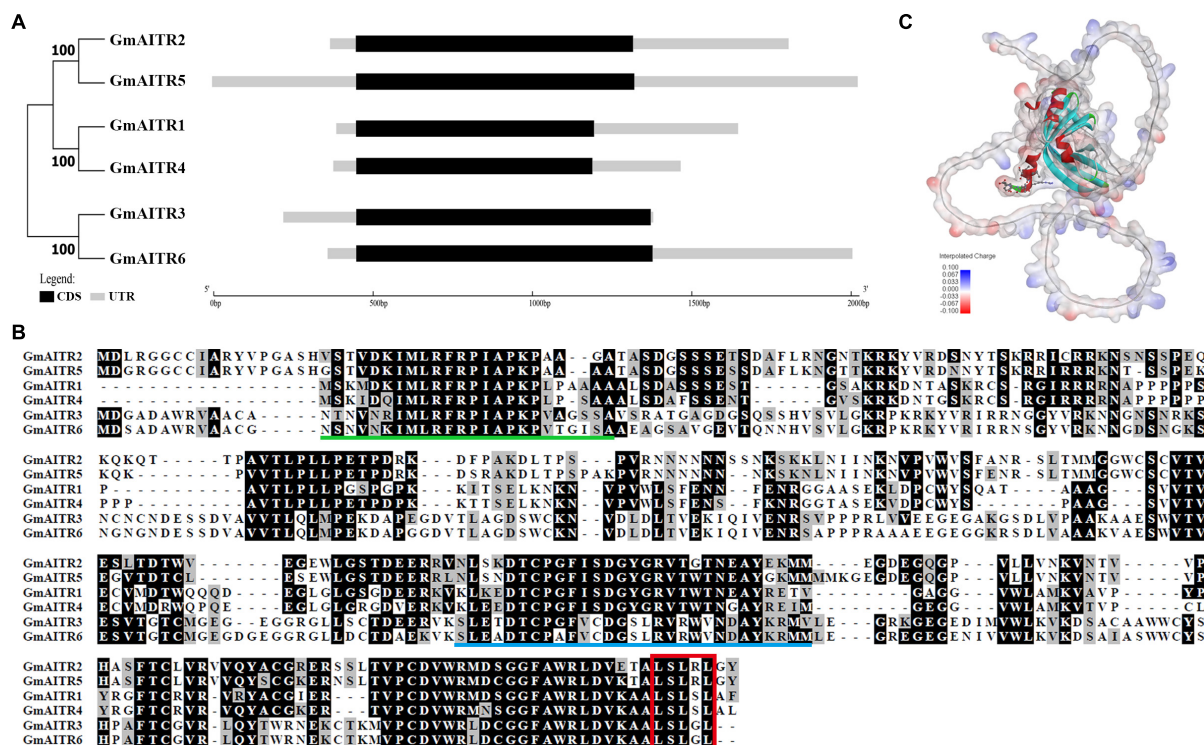


FIGURE 1 | Absciscic acid (ABA) induced transcription repressors (AITRs) in Soybean. **(A)** Phylogenetic relationship and gene structures of *GmAITRs*. Coding sequence (CDS) and untranslated region (UTR) sequences were indicated in black and gray boxes, respectively, and the lengths were drawn to scale. The phylogenetic analysis was performed by using MEGA 7.0. **(B)** Amino acid sequence alignment of *GmAITRs*. Full-length amino acid sequences of *GmAITRs* were obtained from phytosome (<https://phytozome-next.jgi.doe.gov/>), and used for alignment on BioEdit. Identical and similar amino acids are shaded in black and gray, respectively. Conserved motifs are predicted by MEME analysis (<http://meme-suite.org>), and indicated by underlines in different colors. Red box indicates the LxLxL transcriptional repression motif. **(C)** Three-dimensional structure of *GmAITR2* predicted by AlphaFold v2.0 (<https://www.alphafold.ebi.ac.uk/>). The LxLxL repression motif was highlighted in scaled-atom form and the atom charge was added as the surface of the protein.

5'-GGATGCACCGGTACATACC(TGG)-3' targets *GmAITR2* and *GmAITR5*, 5'-GGAGGGGTTTGGGGCGGATA(GGG)-3' targets *GmAITR1* and *GmAITR4*, and 5'-GCGTGACAGGCACG TGCATG(GGG)-3' targets *GmAITR3* and *GmAITR6*. The target sequences in construct two are 5'-GTGGTGTTCGT GTGTGACGG(TGG)-3' targets *GmAITR2* and *GmAITR5*, 5'-G AGGTTTCACGTGCAGGGTG(AGG)-3' targets *GmAITR1* and *GmAITR4*, and 5'-GTGAAAGCTGCGCTCAGTTT (GGG)-3' targets *GmAITR3* and *GmAITR6*. The primers used for making the constructs are listed in **Supplementary Table 2**.

Plant Transformation, Transgenic Plant Selection, and Transgene-Free Mutant Isolation

pYL-CRISPR/Cas9P_{ubi}-B constructs for *GmAITRs* were transformed into the *Agrobacterium tumefaciens* strain of EHA105, and then used to transform soybean by using *Agrobacterium*-mediated cotyledonary node transformation method as previously described (Paz et al., 2004).

Transgenic plants generated were initially examined by using GMO DETECT kit (bar/pat) (Artron Laboratory Inc., Beijing) following the manufacturer's instructions, and then

examined by PCR amplification of *Cas9* gene fragment. Gene editing status in the confirmed T1 transgenic plants was examined by amplifying and sequencing the genomic sequence of *GmAITR* genes. Transgene-free homozygous mutants were isolated from T2 progeny of gene edited T1 plants by PCR amplification of *Cas9* gene fragment, and sequencing of *GmAITR* genes.

DNA Isolation and PCR

DNA was isolated from leaves of the T1 transgenic plants and T2 progeny of gene edited T1 plants by using a method described previously (Edwards et al., 1991).

To confirm the transgenic status of the T1 plants and to isolate transgene-free mutants in T2 progeny of gene edited T1 plants, DNA isolated was used as templates to amplify *Cas9* gene fragment by PCR. The primers used are 5'-CGCTCAGATTGGAGATCAGT-3', and 5'-CGAAGTT CCAAGGGGTGATA-3'.

To examine genome editing status of *GmAITR* genes, DNA isolated was used as templates to amplify genome sequence of *GmAITR* genes by PCR, and the PCR products was isolated and sequenced. The sequencing results were aligned with wild type sequences of the Corresponding *GmAITR* gene. The primers

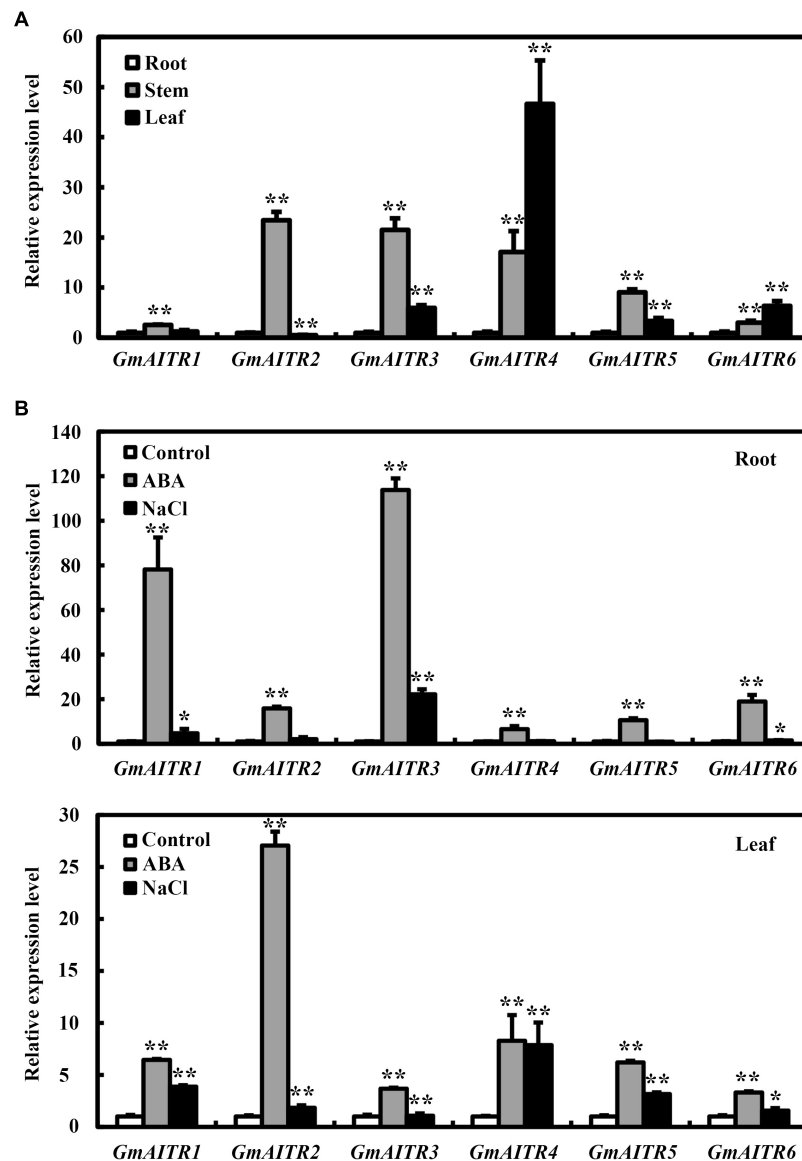


FIGURE 2 | Expression of *GmAITRs* in different tissues and organs, and in response to ABA and salt treatments. **(A)** Expression of *GmAITRs* in different tissues and organs. Roots, stems and leaves were collected from 28-day-old soil-grown plants with the trifoliate leaf fully opened, total RNA was isolated and qRT-PCR was used to examine the expression of *GmAITRs*. The expression of *GmEF-1α* was used as an inner control. The expression levels of *GmAITRs* in roots were set as 1. Data represent the mean \pm SD of three replicates. **(B)** Expression of *GmAITRs* in response to ABA and salt treatments in roots (up panel) and leaves (low panel). Fourteen-day-old seedlings grown in plastic growth bags were exposed to distilled water, 100 μ M ABA or 200 mM NaCl for 6 h, then the roots and leaves were dissected, total RNA were isolated and qRT-PCR was used to examine the expression of *GmAITRs*. The expression of *GmEF-1α* was used as an inner control. The expression levels of *GmAITRs* in distilled water control were set as 1. Data represent the mean \pm SD of three replicates. The experiments were repeated three times with similar results. The asterisks in the figure indicate significant different from the control (* $P < 0.05$; ** $P < 0.01$).

used for PCR amplification of *GmAITR* genes are listed in **Supplementary Table 3**.

Plasmid Isolation, Protoplast Isolation, and Transient Transfection

Plasmids of the reporter and effector constructs were extracted using a GoldHi EndoFree Plasmid Maxi Kit (CWBI) according to the manufacturer's instructions. Protoplasts were isolated

and transfected by following a procedure previously described (Xiong et al., 2019). Briefly, protoplasts were isolated from trifoliate leaves of 2-week-old soil pot-grown Wm82 wild-type plants, plasmids were transfected or co-transfected into the protoplasts isolated, and transfected protoplasts were incubated under darkness at room temperature. For subcellular localization assays, the transfected protoplasts were incubated for 16–18 h, and then GFP fluorescence was examined under an Olympus BX61 fluorescence microscope. For transcription activity assays,

the transfected protoplasts were incubated for 22–24 h, and then GUS activities were measured by using a SynergyTM HT fluorescence microplate reader (BioTEK).

Seed Germination Assays

Healthy and uniform-sized seeds of the Wm82 wild type and the *gmaitr* mutant plants were placed in Petri plates on the surface of two layers of filter papers soaked with 100 μ M ABA 200 mM NaCl, or distilled water as a control. The plates were kept in a growth room, and germinated seeds were count at indicated time points. Each plate contains ten seeds and seeds with radicles longer than 0.5 cm were calculated as germinated seeds at the eleven time points (Kan et al., 2015).

Seedling Growth Assays

Healthy and uniform-sized seeds of the Wm82 wild type and the *gmaitr* mutant plants were germinated and grown with distilled water in plastic growth bags (PhytoTC, Beijing) (Li et al., 2019) for 3 days, and then initiated the salt treatment by adding the 200 mM NaCl solution or fresh distilled water as a control. Two parallel bags were used for each treatment, and two plants for each genotype were included in one bag, and different genotypes in different growth bags were placed in different order to minimize the position effects. After grown in a growth room for 2 weeks, seedlings were taken out from the growth bags for measurement of the shoot and root length.

Field Production Assays

For the agronomic traits comparison, seeds of the Wm82 wild type and the *gmaitr* mutant plants were sown in the experimental fields in plots by genotypes. Each plot in the two normal soil fields includes four rows, and each plot in the saline-alkali field includes three rows. The plot length was 2 m, the space between rows was 0.5 m, and the space between plants in the rows was 10 cm. The seeds were sown in May and the plants were harvested in October in the year 2020.

Statistical Analysis

A statistical analysis of the phenotypic data and expression levels was performed using two-tailed Student's *t*-test in Excel (**P* < 0.05, ***P* < 0.01).

RESULTS

Abscisic Acid Induced Transcription Repressors in Soybean

We have previously identified that there are six genes in soybean encoding AITRs, a number identical to that in Arabidopsis (Tian et al., 2017). Similar to the Arabidopsis *AITR* genes (Tian et al., 2017), all the 6 *GmAITRs* are genes with a single exon (Figure 1A). Phylogenetic analysis shows that GmAITR2 is closely related to GmAITR5, whereas GmAITR1 is closely related to GmAITR4, and together, these four GmAITRs formed one clade. On the other hand, GmAITR3 is closely related to GmAITR6, and they formed another

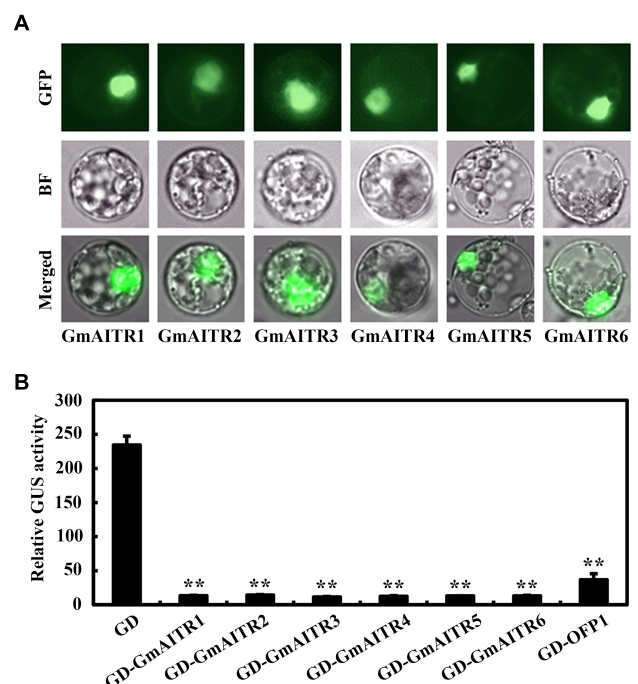


FIGURE 3 | Subcellular localization and transcriptional activities of GmAITRs. **(A)** Subcellular localization of GmAITRs. Plasmid DNA of the 35S::GFP-*GmAITRs* constructs was transiently transfected into protoplasts isolated from unifoliate leaves collected from 10-day-old soybean seedlings, and transfected protoplasts were incubated at room temperature and in darkness for 16–18 h. GFP fluorescence was visualized and pictures were taken under a fluorescence microscope. Up panel, GFP channel images, middle panel, bright-field (BF) images, low panel, merged images. **(B)** Transcriptional activities of GmAITRs. Plasmid DNA of the *GD-GmAITRs* constructs were co-transfected with the effector construct *LD-VP* and the reporter construct *LexA-Gal4::GUS* into protoplasts isolated from unifoliate leaves collected from 10-day-old soybean seedlings, and the transfected protoplasts were incubated at room temperature and in darkness for 22–24 h before GUS activity was assayed. Cotransfection of *GD* and *GD-AtOPF1* were used as negative and positive controls, respectively. Data represent the mean \pm SD of three replicates. The experiments were repeated three times with similar results. **Significant different from the GD control (*P* < 0.01).

clade (Figure 1A). Expanded phylogenetic analysis with AITRs from the dicot plant Arabidopsis, soybean and Medicago and the monocot plant rice (*Oryza sativa*) shows that the two OsAITRs formed a distinct clade, whereas two other clades were formed by AITRs from the three dicot plants, and both of the clades contain AITRs from all the three dicot plants (Supplementary Figure 1).

Sequence alignment shows that GmAITRs shared high amino acid identity and similarity, and contain a conserved LxLxL motif at their C-terminal (Figure 1B). Protein domain assays indicates that these three conserved domains in all the GmAITRs, one at the N-terminal, one in the middle region and the third is the LxLxL motif containing domain at the C-terminal (Figure 1B and Supplementary Figure 2). In addition, GmAITRs are hydrophilic⁶ and non-transmembrane

⁶http://www.detaibio.com/sms2/protein_gravy.html

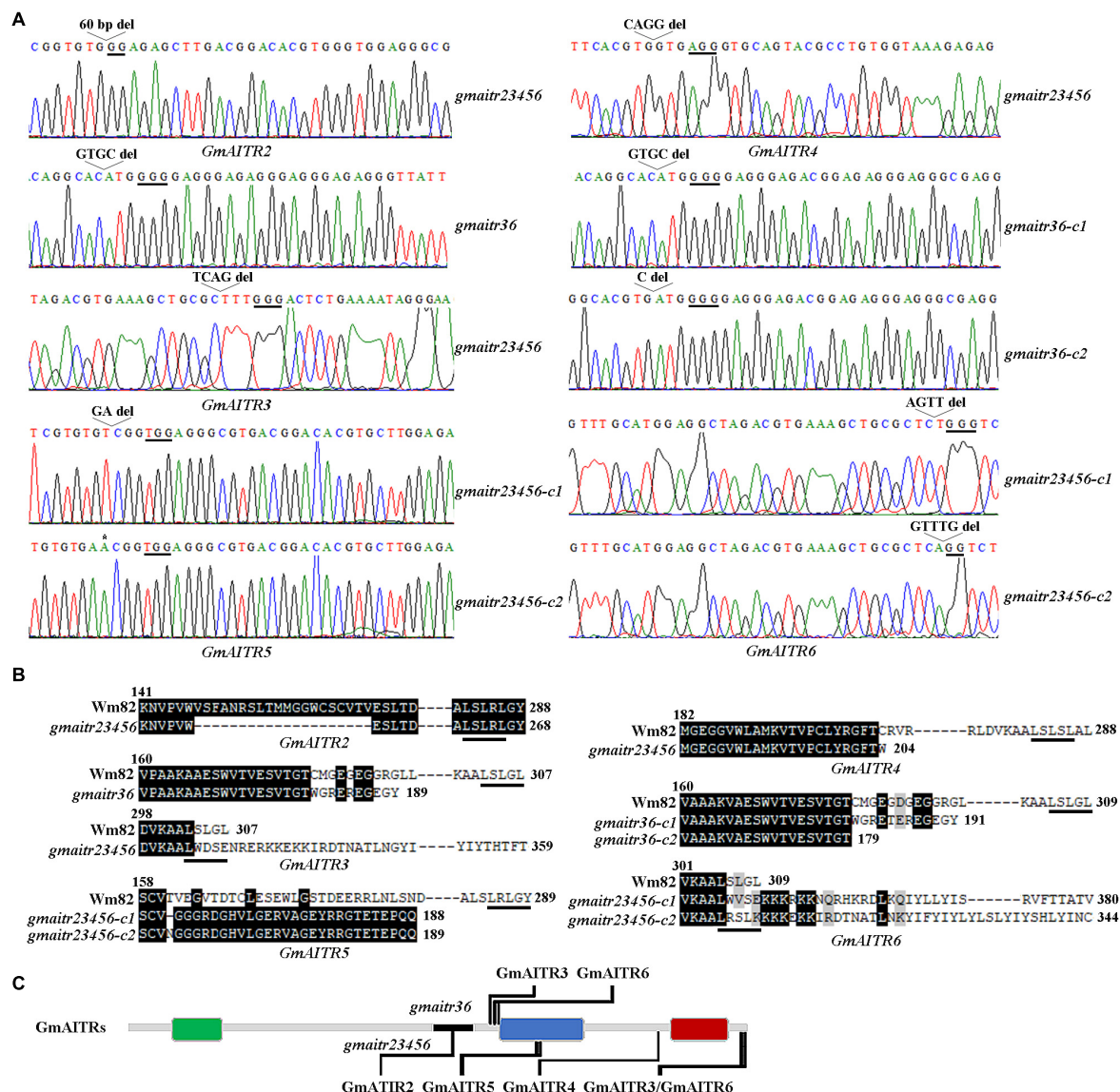


FIGURE 4 | Generation of the *gmaitr36* double and the *gmaitr23456* quintuple mutants. **(A)** Editing status of *GmAITR* genes in the *gmaitr36* double and the *gmaitr23456* quintuple mutants. The mutants were generated by transforming Wm82 wild type soybean with the two CRISPR/Cas9 constructs. DNA was isolated from T1 plants and sequenced to identify gene edited mutants, transgene-free progeny of the edited T1 plants was sequenced to identify homozygous mutants. Underlines indicate the PAM sites. Open arrow heads indicate nucleotide deletions, and asterisk indicates nucleotide insertion. **(B)** Amino acid sequence alignments of GmAITRs in the Wm82 wild type and the *gmaitr36* double and the *gmaitr23456* quintuple mutant plants. Numbers above the sequences indicate the amino acid positions of corresponding GmAITRs in the Wm82 wild type plants, numbers at the end of the sequence indicate the total numbers of amino acids of the corresponding GmAITRs in the Wm82 wild type or the *gmaitr* mutants, and underlines indicate the LxLxL transcriptional repression motifs in the corresponding GmAITRs. **(C)** Schematic diagram of the positions in GmAITRs where amino acids were altered in the *gmaitr36* double and the *gmaitr23456* quintuple mutants. Colored boxes indicate the conserved motifs of GmAITRs predicted by using MEME (<http://meme-suite.org>), black line in the gray line indicate the deletion of 20 amino acids of *GmAITR2* in the *gmaitr23456* quintuple mutants.

proteins⁷, and protein structure prediction with AlphaFold v2.0 (Jumper et al., 2021) indicate that all the GmAITRs have similar three-dimensional structures (Figure 1C and Supplementary Figure 3). These results suggest that GmAITRs may have similar functions.

⁷<http://www.cbs.dtu.dk/services/TMHMM/>

Expression of *GmAITRs* Is Induced by Abscisic Acid and Salt, and *GmAITRs* Function as Transcription Repressors

To examine the functions of GmAITRs in ABA signaling and abiotic stress tolerance, we first examined the expression pattern of *GmAITR* genes. We found that *GmAITR* genes showed diverse expression patterns in the tissues and organs examined. In

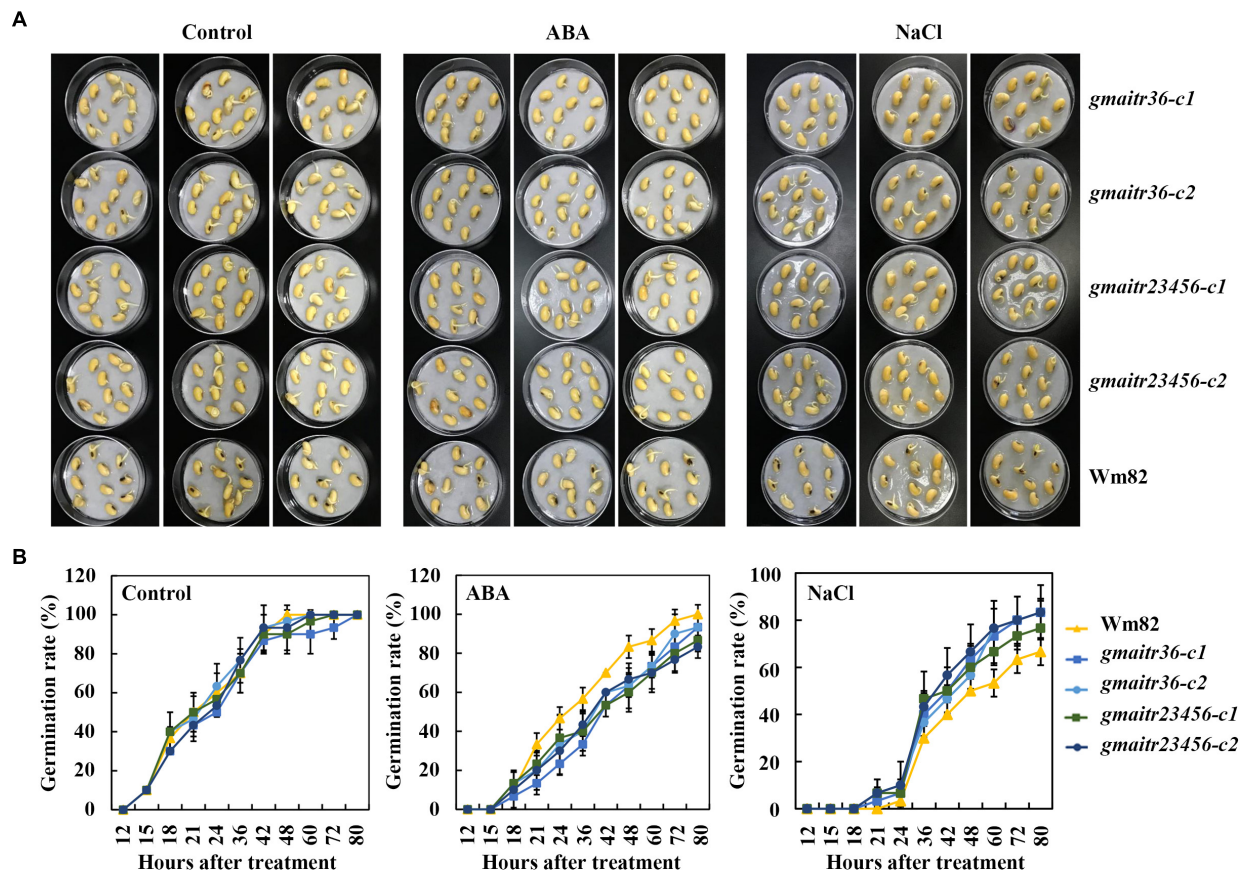


FIGURE 5 | Seed germination of the Wm82 wild-type and the *gmaitr* mutants under ABA and salt treatments. **(A)** Germination of the Wm82 wild-type and the *gmaitr* mutants seeds in plates containing ABA and salt solutions. Healthy and uniform sized seeds of the Wm82 wild-type and the *gmaitr* mutants were incubated on two layers of filter papers in Petri dishes, treated with 100 μ M ABA, 200 mM NaCl, or distilled water as a control. Ten seeds were from each genotype were used for the treatment and every treatment contains three replicates. The pictures were taken 48 h after treatment. **(B)** Seed germination rates the Wm82 wild-type and the *gmaitr* mutant seeds in plates containing ABA and salt solutions. Seeds germinated were counted at indicated time points and seed germination rates were calculated. Data represent the mean \pm SD of three replicates. The experiments were repeated three times with similar results.

general, relative higher expression levels for all the 6 *GmAITSR* genes were observed in stems, and all but *GmAITSR2* also have relative higher expression levels in leaves (Figure 2A). However, difference in expression levels in different tissues and organs were observed for different *GmAITSR* genes, for example, the highest expression level of *GmAITSR1* was observed in stems, but it was only about 2.5-fold of that in root, whereas that of *GmAITSR4* in leaves was nearly 50-fold of that in root (Figure 2A).

We have previously shown that the expression of *GmAITSR* genes is induced by treating excised soybean roots with ABA (Tian et al., 2017). Having shown that *GmAITSRs* showed different expression patterns in the tissues and organs, we then compared ABA response of *GmAITSR* genes in roots and leaves. We found that the expression levels of all the *GmAITSR* genes were increased in response to ABA treatment in both root and leaves, but to different levels. For instance, an ~80- and 110-fold increase for *GmAITSR1* and *GmAITSR3*, respectively in roots, and an ~27-fold increase for *GmAITSR2* in leaves (Figure 2B).

We also examined the expression of *GmAITSRs* in response to salt stress, and found that salt treatment induced the

expression of different *GmAITSR* genes at least in roots or leaves, although to a relative lower levels when compared to ABA treatment (Figure 2B).

We further examined subcellular localization and transcriptional activity of *GmAITSRs* in soybean protoplasts. Similar to the results observed in transfected Arabidopsis protoplasts (Tian et al., 2017), *GmAITSRs* were localized in nucleus (Figure 3A), and they repressed the expressed *Gal4:GUS* reporter gene when recruited to the *Gal4* promoter by the fused *Gal4* DNA binding domain (Figure 3B). These results suggest that *GmAITSRs* function as transcription repressors in soybean.

Generation of Genome Edited Transgene-Free Mutants for *GmAITSR* Genes

Our previous studies have shown that *AITSRs* are conserved in angiosperms, and *AITSR* genes may be good targets for CRISPR/Cas9 genome editing to improve abiotic stress tolerance in crops (Tian et al., 2017; Chen et al., 2021; Wang et al., 2021).

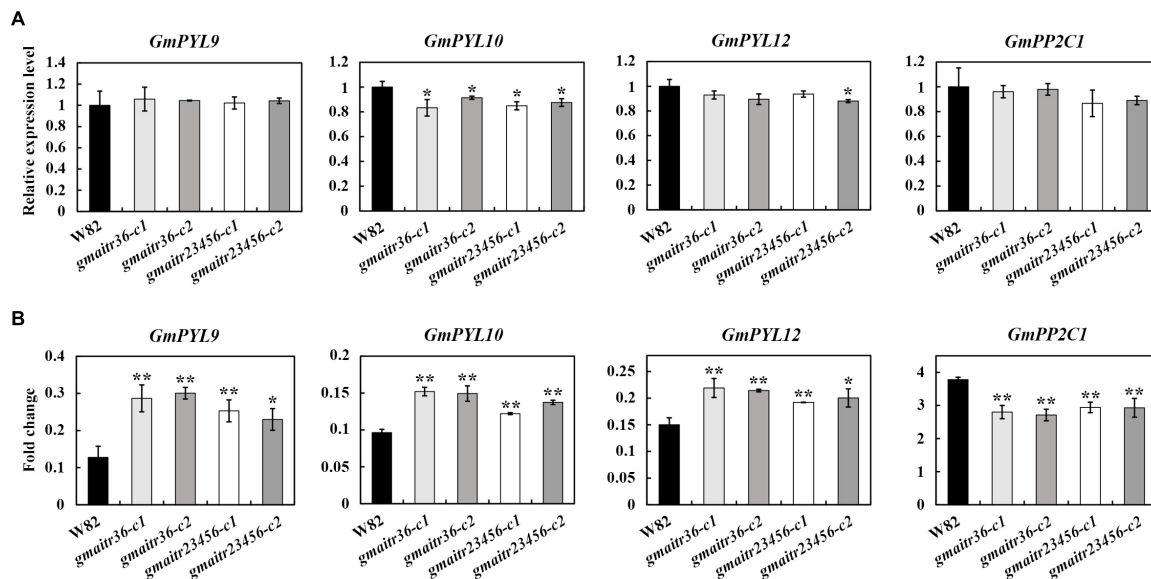


FIGURE 6 | ABA response of the ABA signaling key regulator genes in the Wm82 wild type and the *gmaitr* mutant plants. **(A)** Basal expression levels of ABA signaling key regulator genes in the Wm82 wild type and the *gmaitr* mutant plants. RNA was isolated from 14-day-old seedlings grown in plastic growth bags, total RNA was isolated and qRT-PCR was used to examine the expression of ABA signaling key regulator genes. The expression of *GmEF-1α* was used as an inner control. The expression levels of the corresponding genes in the Wm82 wild type were set as 1. Data represent the mean \pm SD of three replicates. **(B)** ABA responses of the ABA signaling key regulator genes in the Wm82 wild type and the *gmaitr* mutant plants. Fourteen-day-old seedlings of the Wm82 wild type and the *gmaitr* mutant plants grown in plastic growth bags were treated with 100 μ M ABA or solvent as control for 6 h, roots were collected, total RNA was isolated and qRT-PCR was used to examine the expression of ABA signaling key regulator genes. The expression of *GmEF-1α* was used as an inner control. Fold changes were calculated by comparing the expression levels of the corresponding genes in ABA-treated and control seedlings. Data represent the mean \pm SD of three replicates. The experiments were repeated three times with similar results. The asterisks indicate significant differences (* P < 0.05, ** P < 0.01).

Our results described above indicate that GmAITRs and Arabidopsis AITRs shared similar features, we therefore decided to generate transgene-free mutants of *GmAITR* genes by using CRISPR/Cas9 genome editing, and examine their response to ABA and abiotic stresses.

Two different CRISPR/Cas9 constructs were generated by using the *pYL-CRISPR/Cas9_{ubi}-B* vector (Ma et al., 2015), and each construct contains three target sequences with each is aimed to target a pair of *GmAITR* genes. The Wm82 wild type soybean was used for plant transformation, and gene edited status were examined in T1 plants, and transgene-free homozygous mutants were isolated from progeny of gene edited T1 plants. Editing of *GmAITR3* and *GmAITR6* were observed in T1 plants generated with one construct, and editing of *GmAITR2-GmAITR6* were observed in T1 plants generated with another construct. Finally, transgene-free *gmaitr3 gmaitr6* (*gmaitr36*) double and *gmaitr2 gmaitr3 gmaitr4 gmaitr5 gmaitr6* (*gmaitr23456*) quintuple homozygous mutants were obtained from construct one and two transformed plants, respectively.

In all the mutants obtained, either a single nucleotide insertion or one to up to 60 nucleotides deletion was occurred at the target sites for the *GmAITR* genes (Figure 4A), resulting in changes of amino acid sequence of the corresponding GmAITR proteins. In both *gmaitr36* double mutants, amino acid substitutions and premature stop occurred in GmAITR3, whereas amino acid substitutions and premature stop occurred in GmAITR6 in the *gmaitr36-c1* double mutant, and immediately premature

stop occurred in GmAITR6 in the *gmaitr36-c2* double mutant, respectively (Figure 4B). In the *gmaitr23456* quintuple mutants, 20 amino acids deletion occurred in GmAITR2, an amino acid substitution and premature stop occurred in GmAITR4, and amino acid substitutions and premature stop occurred in GmAITR5 (Figure 4B). However, nucleotides deletions in *GmAITR3* and *GmAITR6* in the *gmaitr23456* quintuple mutants led to amino acid substitution and addition of extra amino acids in corresponding GmAITR proteins (Figure 4B). The positions of amino acids changes in the GmAITR proteins for the *gmaitr36* double and *gmaitr23456* quintuple mutants were diagrammed in Figure 4C. Moreover, protein structures of genome edited GmAITRs were predicted by AlphaFold v2.0 (Supplementary Figure 3), and obvious differences can be found in *gmaitr36* double mutants for protein GmAITR3 and GmAITR6. In *gmaitr23456* quintuple mutants, protein structures of GmAITR4 and GmAITR5 were severely damaged compared with wild type, while GmAITR2, GmAITR3, and GmAITR6 preserved similar structures as wild type.

The *gmaitr* Mutants Are Hypersensitivity to Absciscic Acid

By using seed germination assays, we examined ABA response of the *gmaitr* mutants generated. Different from the results observed in the Arabidopsis *aitr* mutants, which showed a decreased ABA sensitivity (Tian et al., 2017; Chen et al., 2021), we found

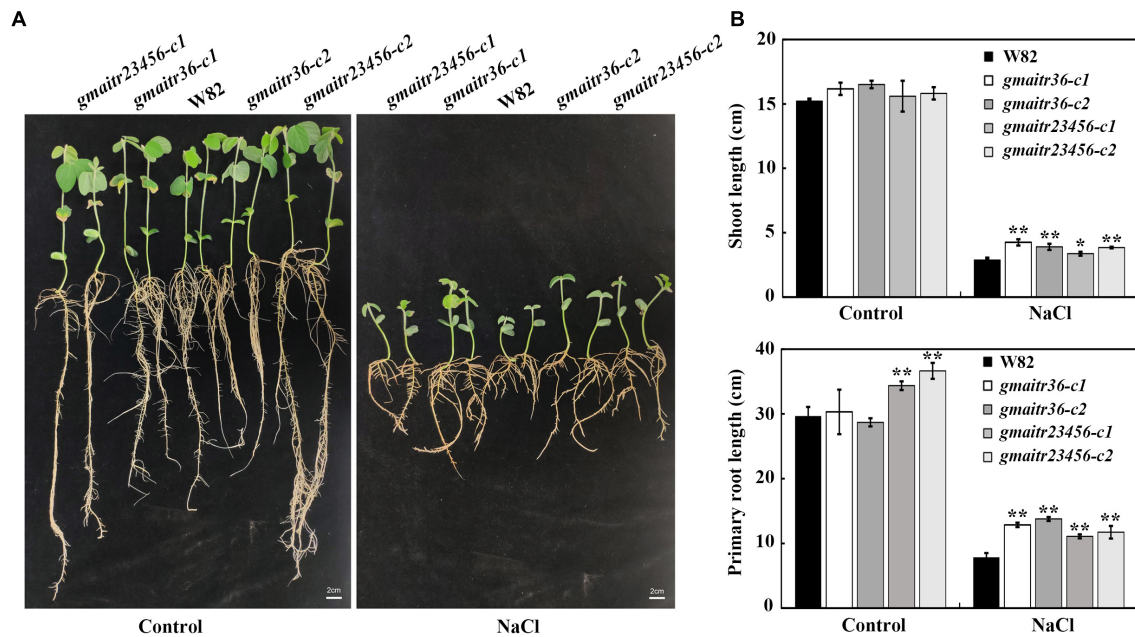


FIGURE 7 | Salt tolerance of the *gmaitr* mutants. **(A)** Seedlings of the Wm82 wild type plant and the *gmaitr* mutants in response to salt treatment. Seeds were germinated and grown in distilled water-containing plastic growth bags for 3 days in a growth chamber, and then the distilled water was replaced by 200 mM NaCl or fresh distilled water as a control. For each treatment, seedlings were grown in two growth bags with two plants for each genotype. Two weeks after the treatment, representative seedlings were taken out from the growth bags and images were taken using a digital camera. Bar, 2 cm. **(B)** Shoot and primary root length of the control and salt treated seedlings of Wm82 wild type and the *gmaitr* mutant plants. Length of the shoot and primary root length of the Wm82 wild type and the *gmaitr* mutant seedlings were measured 2 weeks after the treatment. Data represent the mean \pm SD of four seedlings. The asterisks indicate significant differences (* $P < 0.05$, ** $P < 0.01$).

that seeds of all the *gmaitr* mutants were more sensitivity to ABA treatment when compared to the Wm82 wild type seeds (Figure 5A). Quantitative assays show that no difference was observed for the Wm82 wild type and the mutant seeds on control plates (plates soaked with distilled water, which is the dissolvent of ABA and salt solution), seeds of all the plants reached a maximum germinate rate, i.e., $\sim 100\%$ 48 h after treatment. On the other hand, when compared to the Wm82 wild type seeds, a reduced germination rate was observed for seeds of all the mutants on the ABA treated plates (Figure 5B), indicating that ABA sensitivity in the mutants was increased. However, we found that germination rate of the *gmaitr36* double mutant seeds is largely indistinguishable from that of the *gmaitr23456* quintuple mutant seeds (Figure 5B).

Our previously results indicated that AITRs in Arabidopsis function as feedback regulators in ABA signaling by inhibiting ABA responses of some ABA signaling regulators genes (Tian et al., 2017; Chen et al., 2021). Having shown that ABA response in the *gmaitr* mutants was affected, we further examined if expression levels of the core ABA signaling regulator genes may be changed in the *gmaitr* mutants. We treated the *gmaitr* mutants and Wm82 seedlings with different concentration of ABA solution, and ABA key regulator genes were significantly induced in soybean seedlings treated with 100 μ M ABA, thus 100 μ M ABA was used for expression analysis. We found that the basal expression levels of some ABA signaling key regulator genes identified previously (Bai et al., 2013), including the GmPYL

receptor genes GmPYL9, GmPYL10, GmPYL12, and the PP2C phosphatase gene GmPP2C1 remained largely unchanged in the *gmaitr* mutants (Figure 6A). However, ABA induced responses of these genes were reduced in the *gmaitr* mutants, even though little, if any difference was observed between the *gmaitr36* double and the *gmaitr23456* quintuple mutants (Figure 6B).

The *gmaitr* Mutant Plants Are Tolerant to Salt Stress

Changes in the expression levels of the ABA signaling regulator genes including Arabidopsis AITR genes have been shown to affect plant abiotic stress tolerance (Fujita et al., 2009; Park et al., 2015; Yoshida et al., 2015; Zhao et al., 2016; Tian et al., 2017; Chen et al., 2021), but so far only *aitr* mutants showed enhanced tolerance to drought and salt, make AITRs good candidate genes for CRISPR/Cas9 genome editing to improve abiotic stress tolerance in plants (Tian et al., 2017; Chen et al., 2021).

To examine if mutation of GmAITR genes may indeed improve abiotic stress tolerance in soybean, we first examined the effects of salt treatment on seed germination of the *gmaitr* mutants. We found that the *gmaitr* mutant seeds showed enhanced tolerance to salt treatment (Figure 5A), and quantitative assays showed that an increased germination rate were observed for seeds of *gmaitr* mutants at all the time points examined (Figure 5B). But similar to the results observed with ABA treatment, little, if any difference was observed between

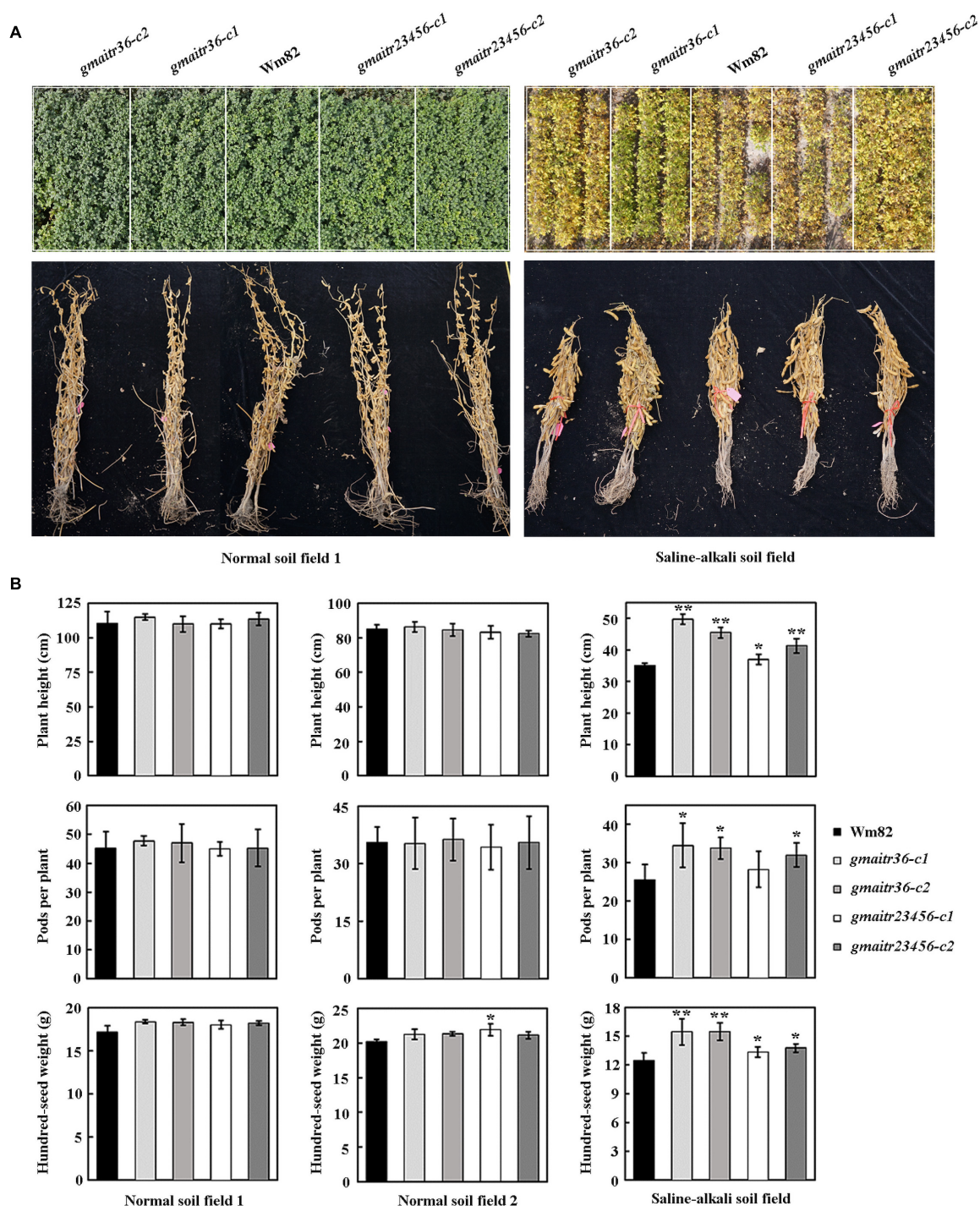


FIGURE 8 | Field production of the Wm82 wild type and the *gmair* mutant plants in normal and saline-alkali soil lands. **(A)** Plants of the Wm82 wild type and the *gmair* mutants in normal and saline-alkali soil lands. The Wm82 wild-type and the *gmair* mutants were grown in two normal soil and one saline-alkali land (pH 8.1–9.8; soluble saline 0.1–0.7%) for field production analysis in the year 2020. Seeds were planted in plots by genotypes. Each plot in normal fields includes four rows, and plots in saline-alkali field include three rows. Numbers of seeds planted in a row for each plot were the same. *Upper panel*, field images of 4-month-old plants from one of the normal soil land the saline-alkali soil land. The white frames were used to indicate the edges of the plots. *Low panel*, images of five bundled representative mature plants for each genotype from one of the normal soil land the saline-alkali soil land. **(B)** Yield indexes of the Wm82 wild type and the *gmair* mutant plants in normal and saline-alkali soil lands. The Wm82 wild-type and the *gmair* mutant plants were harvested and plants randomly selected were used for yield indexes measurement, including plant height, pods produced per plant, and hundred-seed weight. For each field, the measurement was repeated four times with four different set of plants. Each set of plants contain five randomly selected plants from each plot. Data represent the mean \pm SD of at least four replicates. The asterisks indicate significant differences (* $P < 0.05$; ** $P < 0.01$).

the *gmaitr36* double and the *gmaitr23456* quintuple mutant seeds (**Figure 5B**).

We then examined the effects of salt treatment on seedling growth of the *gmaitr* mutants. As shown in **Figure 7A**, the *gmaitr* mutant seedlings showed enhanced tolerance to salt treatment, as they produced longer roots and shoots when compared with the Wm82 wild type seedlings (**Figure 7B**).

At last, we compared growth and yield of the Wm82 wild type and the *gmaitr* mutant plants in both normal soil field and saline-alkali soil field. We found that the *gmaitr* mutant plants are morphological similar to the Wm82 wild type plants in the normal soil field, but growth better in the saline soil field (**Figure 8A**). Both the Wm82 wild type and the *gmaitr* mutant plants reached a height of ~110 cm at mature stage, with ~45 pods per plant, and produced seed with hundred-seed weight of ~18 g (**Figure 8B**). Plants height, number of pods per plants and hundred-seed weight were all dramatically decreased in the saline-alkali soil field, however, the *gmaitr* mutant plants were less affected (**Figure 8B**).

DISCUSSION

Even though CRISPR/Cas9 genome editing has been successfully used to improve important agronomic traits in several different crops (Ma et al., 2015; Gao et al., 2016; Li X. et al., 2017; Lu et al., 2017; Shimatani et al., 2017; He et al., 2018; Zsögön et al., 2018; Chen et al., 2019), identification of suitable candidate genes in ABA signaling pathway for genome editing to improve abiotic stress tolerance in crops is a big challenge. In soybean, several different types of transcription factors involved in abiotic stress tolerance have been reported to be related to ABA signaling pathway, such as the AP2/ERF transcription factor GmERF3 (Zhang et al., 2009), the bZIP transcription factor GmbZIP1, GmbZIP15 and GmFDL19 (Gao et al., 2011; Li Y. et al., 2017; Zhang et al., 2020), the R2R3 MYB transcription factor GmMYB84 (Wang et al., 2017), the WRKY transcription factor GmWRKY12 and GmWRKY54 (Shi et al., 2018; Wei et al., 2019), and the NAC transcription factor GmSIN1, GmNAC06 and GmNAC8 (Li et al., 2019, 2021; Yang et al., 2020). However, among all these transcription factors, only GmbZIP15 functioned as a negative regulator of abiotic stress tolerance in soybean, yet no enhanced tolerance was observed in the transgenic soybean plants expressing a repressor form of GmbZIP15 (Zhang et al., 2020). These results suggest that none of these transcription factor genes can serve as targets for CRISPR/Cas9 gene editing to improve abiotic stress tolerance in soybean.

We have previously identified AITRs as a novel family of transcription factors conserved in angiosperms, and loss-of-function of *AITR* genes enhanced abiotic stress tolerance in Arabidopsis without fitness costs, indicating that AITRs may be good candidates for gene editing to improve abiotic stress tolerance in crops (Tian et al., 2017; Chen et al., 2021). By using of a combination of different assays including gene expression assays, transcriptional activity assays, generation of transgene-free gene edited mutants, and physiological and field yield analysis, we show that GmAITRs are ABA and salt inducible

transcription repressors, and GmAITRs can be targeted to improve salinity stress tolerance in soybean.

First, we show that the expression of GmAITRs was induced by both ABA and salt treatments, even though these genes have different expression pattern, and there are difference among these genes in responses to ABA and salt (**Figure 2**). Second, we found that, similar to the results observed in Arabidopsis protoplasts (Tian et al., 2017), GmAIR proteins localized in nucleus and they repressed reporter gene expression in soybean protoplasts (**Figure 3**). Third, ABA inhibited seed germination was affected in the *gmaitr* mutants (**Figure 5**), and ABA response of some ABA signaling key regulator genes was altered in the *gmaitr* mutants (**Figure 6**). These results suggest that GmAITRs are ABA responsive transcription repressors and they regulate ABA response in soybean via affecting ABA signaling. Forth, the *gmaitr* mutants showed enhanced tolerance to salt in both seed germination and seedling growth assays (**Figures 5, 7**). Last but not least, field experiments suggest that the *gmaitr* mutants performed better in the saline-alkali soils when compared to the Wm82 wild type plants (**Figure 8**). These results suggest that genome editing of GmAIR genes is able to enhance salt tolerance in soybean.

It should be noted that in ABA inhibited seed germination assays, the *gmaitr* mutants showed increased sensitivity to ABA (**Figure 5**), a result different from that of the Arabidopsis *aitr* mutants, which showed decreased sensitivity to ABA (Tian et al., 2017; Chen et al., 2021), suggest that there is some difference between GmAITRs and Arabidopsis AITRs in regulating ABA responses. However, the *gmaitr* mutants also showed enhanced tolerance to salt (**Figures 5, 7**), similar to that observed in the Arabidopsis *aitr* mutants (Tian et al., 2017; Chen et al., 2021), making them good targets for genome editing to improve abiotic stress tolerance in soybean.

We also noted that the *gmaitr23456* quintuple mutants are largely indistinguishable to the *gmaitr36* double mutants in both ABA and salt tolerance assays, and in field growth conditions (**Figures 5, 7, 8**). Even though we cannot rule out the possibility that some of the GmAITRs may have a dominate roles in regulation ABA response and salt tolerance, as we previously observed for the Arabidopsis AITRs (Chen et al., 2021). Based on the conserved motif analysis (**Figure 4**) and protein structure prediction results (**Supplementary Figure 3**), a possible explanation is that the editing to *GmAIR2*, *GmAIR3*, and *GmAIR6* in the *gmaitr23456* quintuple mutants may not led to loss-of-function of these genes. First, as the genome editing of *GmAIR2* in the *gmaitr23456* quintuple mutants only resulted in a deletion of 20 amino acids outside the conserved motifs (**Figure 4**), whereas genome editing of both *GmAIR3* and *GmAIR6* in the *gmaitr23456* quintuple mutants only disrupted the LxLxL motif at the C-terminal of *GmAIR3* and *GmAIR6*, respectively. Our previously results with Arabidopsis AITRs have already shown that the deletion of LxLxL motif affected AITRs' transcriptional repression activities, but they are still able to function as transcription repressors (Tian et al., 2017). Second, according to the three-dimensional protein structure prediction, the protein binding pockets structure, which is important for protein functionality (Stank et al., 2016), were

barely not damaged for GmAITR2, GmAITR3 and GmAITR6 in the *gmaitr23456* quintuple mutants compared with wild type (**Supplementary Figure 3**). Therefore, it will be of great interest to generate high-order loss-of-function mutants of *GmAITR* genes and to examine if increased tolerance to abiotic stresses can be achieved, and if there are any fitness costs. It will be of great interest to compare physiological/biochemical index in the Wm82 and the *gmaitr* mutants, and use more negative controls for the ABA and salt related response analysis, therefore to understand the subtle changes and physiological mechanism of GmAITR in abiotic stress tolerance. It will be also of great interest to edit *GmAITR* genes in soybean cultivars with other good agronomic traits to see if enhanced abiotic stress tolerance can be obtained without affecting these agronomic traits, thereby accelerating the molecular breeding process of soybean with different benefit agronomic traits.

On the other hand, considering that in all the major crops, AITRs are encoded by multiple genes (Tian et al., 2017), loss-of-function of a few *AITR* genes can already led to enhanced abiotic stress tolerance making it more practicable for editing *AITR* genes to improve abiotic stress tolerance in crops. After all, it is not easy to edit all the *AITR* genes simultaneously in a crop.

Nevertheless, our results show that GmAITRs are involved in the regulation of ABA response and abiotic stress tolerance in soybean, and CRISPR/Cas9 genome editing of *GmAITR* genes is able to enhance salt tolerance in soybean.

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.

REFERENCES

- Bai, G., Yang, D., Zhao, Y., Ha, S., Yang, F., Ma, J., et al. (2013). Interactions between soybean ABA receptors and type 2C protein phosphatases. *Plant Mol. Biol.* 83, 651–664. doi: 10.1007/s11103-013-0114-4
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208.
- Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* 70, 667–697.
- Chen, S., Zhang, N., Zhou, G., Hussain, S., Ahmed, S., Tian, H., et al. (2021). Knockout of the entire family of AITR genes in *Arabidopsis* leads to enhanced drought and salinity tolerance without fitness costs. *BMC Plant Biol.* 21:137. doi: 10.1186/s12870-021-02907-9
- Deshmukh, R., Sonah, H., Patil, G., Chen, W., Prince, S., Mutava, R., et al. (2014). Integrating omic approaches for abiotic stress tolerance in soybean. *Front. Plant Sci.* 5:244. doi: 10.3389/fpls.2014.00244
- Dong, T., Park, Y., and Hwang, I. (2015). Abscisic acid: biosynthesis, inactivation, homeostasis and signalling. *Essays Biochem.* 58, 29–48. doi: 10.1042/bse0580029
- 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.

AUTHOR CONTRIBUTIONS

SW, BL, YD, XF, and JP conceived the study. TW and SW designed the experiments, analyzed the data, and drafted the manuscript. HT made the CRISPR/Cas9 constructs. HX and XQ generated the mutants. TW, WW, and GL examined the gene editing status in the mutants. TW, XD, SH, YL, YC, CW, and RL did the experiments. QD performed the bioinformatics analysis. All authors participated in the revision of the manuscript, read and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (32071938), the National Key R&D Program of China (2016YFD0101900), and a startup funding from Linyi University (LYDX2019BS039). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Yaoguang Liu (South China Agricultural University) for sharing the *pYL-CRISPR/Cas9P_{ubi}-B* vector, and all the lab members for helpful discussion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.779598/full#supplementary-material>

- Fujii, H., Verslues, P. E., and Zhu, J. K. (2007). Identification of two protein kinases required for abscisic acid regulation of seed germination, root growth, and gene expression in *Arabidopsis*. *Plant Cell* 19, 485–494. doi: 10.1105/tpc.106.04.8538
- Fujii, H., and Zhu, J. K. (2009). *Arabidopsis* mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. *Proc. Natl. Acad. Sci. U. S. A.* 106, 8380–8385. doi: 10.1073/pnas.0903144106
- Fujita, M., Fujita, Y., Noutoshi, Y., Takahashi, F., Narusaka, Y., Yamaguchi-Shinozaki, K., et al. (2006). Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. *Curr. Opin. Plant Biol.* 9, 436–442. doi: 10.1016/j.pbi.2006.05.014
- Fujita, Y., Nakashima, K., Yoshida, T., Katagiri, T., Kidokoro, S., Kanamori, N., et al. (2009). Three SnRK2 protein kinases are the main positive regulators of abscisic acid signaling in response to water stress in *Arabidopsis*. *Plant Cell Physiol.* 50, 2123–2132. doi: 10.1093/pcp/pcp147
- Gao, S. Q., Chen, M., Xu, Z. S., Zhao, C. P., Li, L., Xu, H. J., et al. (2011). The soybean GmbZIP1 transcription factor enhances multiple abiotic stress tolerances in transgenic plants. *Plant Mol. Biol.* 75, 537–553. doi: 10.1007/s11103-011-9738-4
- Gao, X., Chen, J., Dai, X., Zhang, D., and Zhao, Y. (2016). An effective strategy for reliably isolating heritable and Cas9-free *Arabidopsis* mutants generated by RISPR/Cas9-mediated genome editing. *Plant Physiol.* 171, 1794–1800. doi: 10.1104/pp.16.00663

- Gosti, F., Beaudoin, N., Serizet, C., Webb, A. A., Vartanian, N., and Giraudat, J. (1999). ABI1 protein phosphatase 2C is a negative regulator of abscisic acid signaling. *Plant Cell* 11, 1897–1910. doi: 10.2307/3871085
- Guo, J., Yang, X., Weston, D. J., and Chen, J. G. (2011). Abscisic acid receptors: past, present and future. *J. Integr. Plant Biol.* 53, 469–479. doi: 10.1111/j.1744-7909.2011.01044.x
- He, Y., Zhu, M., Wang, L., Wu, J., Wang, Q., Wang, R., et al. (2018). Programmed self-elimination of the CRISPR/Cas9 construct greatly accelerates the isolation of edited and transgene-free rice plants. *Mol. Plant* 11, 1210–1213. doi: 10.1016/j.molp.2018.05.005
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589.
- Kan, G., Zhang, W., Yang, W., Ma, D., Zhang, D., Hao, D., et al. (2015). Association mapping of soybean seed germination under salt stress. *Mol. Genet. Genomics* 290, 2147–2162. doi: 10.1007/s00438-015-1066-y
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054
- Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., et al. (2013). Multiplex and homologous recombination-mediated genome editing in *Arabidopsis* and *Nicotiana benthamiana* using guide RNA and Cas9. *Nat. Biotechnol.* 31, 688–691. doi: 10.1038/nbt.2654
- Li, M., Chen, R., Jiang, Q., Sun, X., Zhang, H., and Hu, Z. (2021). GmNAC06, a NAC domain transcription factor enhances salt stress tolerance in soybean. *Plant Mol. Biol.* 105, 333–345. doi: 10.1007/s11103-020-01091-y
- Li, S., Wang, N., Ji, D., Zhang, W., Wang, Y., Yu, Y., et al. (2019). A GmSIN1/GmNCED3s/GmRbohBs feed-forward loop acts as a signal amplifier that regulates root growth in soybean exposed to salt stress. *Plant Cell* 31, 2107–2130. doi: 10.1105/tpc.18.00662
- Li, X., Xie, Y., Zhu, Q., and Liu, Y. G. (2017). Targeted genome editing in genes and cis-regulatory regions improves qualitative and quantitative traits in crops. *Mol. Plant* 10, 1368–1370. doi: 10.1016/j.molp.2017.10.009
- Li, Y., Chen, Q., Nan, H., Li, X., Lu, S., Zhao, X., et al. (2017). Overexpression of GmFDL19 enhances tolerance to drought and salt stresses in soybean. *PLoS One* 12:e0179554. doi: 10.1371/journal.pone.0179554
- Lu, H. P., Liu, S. M., Xu, S. L., Chen, W. Y., Zhou, X., Tan, Y. Y., et al. (2017). CRISPR-S: an active interference element for a rapid and inexpensive selection of genome-edited, transgene-free rice plants. *Plant Biotechnol. J.* 15, 1371–1373. doi: 10.1111/pbi.12788
- Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., et al. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* 8, 1274–1284. doi: 10.1016/j.molp.2015.04.007
- Manavalan, L. P., Guttikonda, S. K., Tran, L. S., and Nguyen, H. T. (2009). Physiological and molecular approaches to improve drought resistance in soybean. *Plant Cell Physiol.* 50, 1260–1276. doi: 10.1093/pcp/pcp082
- Matres, J. M., Hilscher, J., Datta, A., Armario-Nájera, V., Baysal, C., He, W., et al. (2021). Genome editing in cereal crops: an overview. *Transgenic Res.* 30, 461–498. doi: 10.1007/s11248-021-00259-6
- Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D., and Kamoun, S. (2013). Targeted mutagenesis in the model plant *Nicotiana benthamiana* using Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* 31, 691–693. doi: 10.1038/nbt.2655
- Park, S. Y., Peterson, F. C., Mosquna, A., Yao, J., Volkman, B. F., and Cutler, S. R. (2015). Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* 520, 545–548. doi: 10.1038/nature14123
- Paz, M., Shou, H., Guo, Z., Zhang, Z., Banerjee, A., and Wang, K. (2004). Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. *Euphytica* 136, 167–179. doi: 10.1023/b:euph.0000030670.36730.a4
- 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
- Rodriguez, P. L., Leube, M. P., and Grill, E. (1998). Molecular cloning in *Arabidopsis thaliana* of a new protein phosphatase 2C (PP2C) with homology to ABI1 and ABI2. *Plant Mol. Biol.* 38, 879–883. doi: 10.1023/a:1006012218704
- Rushton, D. L., Tripathi, P., Rabara, R. C., Lin, J., Ringler, P., Boken, A. K., et al. (2012). WRKY transcription factors: key components in abscisic acid signalling. *Plant Biotechnol. J.* 10, 2–11. doi: 10.1111/j.1467-7652.2011.00634.x
- Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* 31, 686–688. doi: 10.1038/nbt.2650
- Shi, W. Y., Du, Y. T., Ma, J., Min, D. H., Jin, L. G., Chen, J., et al. (2018). The WRKY Transcription factor GmWRKY12 confers drought and salt tolerance in Soybean. *Int. J. Mol. Sci.* 19:4087. doi: 10.3390/ijms19124087
- Shimatani, Z., Kashojiya, S., Takayama, M., Terada, R., Arazoe, T., Ishii, H., et al. (2017). Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* 35, 441–443. doi: 10.1038/nbt.3833
- Song, L., Huang, S. C., Wise, A., Castanon, R., Nery, J. R., Chen, H., et al. (2016). A transcription factor hierarchy defines an environmental stress response network. *Science* 354:aag1550. doi: 10.1126/science.aag1550
- Stank, A., Kokh, D. B., Fuller, J. C., and Wade, R. C. (2016). Protein binding pocket dynamics. *Acc. Chem. Res.* 49, 809–815.
- Tian, H., Chen, S., Yang, W., Wang, T., Zheng, K., Wang, Y., et al. (2017). A novel family of transcription factors conserved in angiosperms is required for ABA signalling. *Plant Cell Environ.* 40, 2958–2971. doi: 10.1111/pce.13058
- Tiwari, S. B., Hagen, G., and Guilfoyle, T. J. (2004). Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* 16, 533–543. doi: 10.1105/tpc.017384
- Umezawa, T., Nakashima, K., Miyakawa, T., Kuromori, T., Tanokura, M., Shinozaki, K., et al. (2010). Molecular basis of the core regulatory network in ABA responses: sensing, signaling and transport. *Plant Cell Physiol.* 51, 1821–1839. doi: 10.1093/pcp/pcq156
- Vanliyodan, B., Ye, H., Song, L., Murphy, M., Shannon, J. G., and Nguyen, H. T. (2017). Genetic diversity and genomic strategies for improving drought and waterlogging tolerance in soybeans. *J. Exp. Bot.* 68, 1835–1849. doi: 10.1093/jxb/erw433
- Wang, N., Zhang, W., Qin, M., Li, S., Qiao, M., Liu, Z., et al. (2017). Drought tolerance conferred in soybean (*Glycine max.* L.) by GmMYB84, a Novel R2R3-MYB transcription factor. *Plant Cell Physiol.* 58, 1764–1776. doi: 10.1093/pcp/pcx111
- Wang, S., Tiwari, S. B., Hagen, G., and Guilfoyle, T. J. (2005). AUXIN RESPONSE FACTOR7 restores the expression of auxin-responsive genes in mutant *Arabidopsis* leaf mesophyll protoplasts. *Plant Cell* 17, 1979–1993. doi: 10.1105/tpc.105.031096
- Wang, T., Dong, Q., Wang, W., Chen, S., Cheng, Y., Tian, H., et al. (2021). Evolution of AITR family genes in cotton and their functions in abiotic stress tolerance. *Plant Biol.* 23, 58–68. doi: 10.1111/plb.13218
- Wang, W., Vinocur, B., and Altman, A. (2003). Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* 218, 1–14. doi: 10.1007/s00425-003-1105-5
- Wei, W., Liang, D. W., Bian, X. H., Shen, M., Xiao, J. H., Zhang, W. K., et al. (2019). GmWRKY54 improves drought tolerance through activating genes in abscisic acid and Ca²⁺ signaling pathways in transgenic soybean. *Plant J.* 100, 384–398. doi: 10.1111/tpj.14449
- Xiong, L., Li, C., Li, H., Lyu, X., Zhao, T., Liu, J., et al. (2019). A transient expression system in soybean mesophyll protoplasts reveals the formation of cytoplasmic GmCRY1 photobody-like structures. *Sci. China Life Sci.* 62, 1070–1077. doi: 10.1007/s11427-018-9496-5
- Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., and Prasanna, B. M. (2012). Whole-genome strategies for marker-assisted plant breeding. *Mol. Breeding* 29, 833–854. doi: 10.1007/s11032-012-9699-6
- Yang, C., Huang, Y., Lv, W., Zhang, Y., Bhat, J. A., Kong, J., et al. (2020). GmNAC8 acts as a positive regulator in soybean drought stress. *Plant Sci.* 293:110442. doi: 10.1016/j.plantsci.2020.110442

- Yoshida, T., Fujita, Y., Maruyama, K., Mogami, J., Todaka, D., Shinozaki, K., et al. (2015). Four *Arabidopsis* AREB/ABF transcription factors function predominantly in gene expression downstream of SnRK2 kinases in abscisic acid signaling in response to osmotic stress. *Plant Cell Environ.* 38, 35–49. doi: 10.1111/pce.12351
- Yoshida, T., Mogami, J., and Yamaguchi-Shinozaki, K. (2014). ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Curr. Opin. Plant Biol.* 21, 133–139.
- Zhang, G., Chen, M., Li, L., Xu, Z., Chen, X., Guo, J., et al. (2009). Overexpression of the soybean GmERF3 gene, an AP2/ERF type transcription factor for increased tolerances to salt, drought, and diseases in transgenic tobacco. *J. Exp. Bot.* 60, 3781–3796. doi: 10.1093/jxb/erp214
- Zhang, M., Liu, Y., Cai, H., Guo, M., Chai, M., She, Z., et al. (2020). The bZIP transcription factor GmbZIP15 negatively regulates salt- and drought-stress responses in soybean. *Int. J. Mol. Sci.* 21:7778. doi: 10.3390/ijms21207778
- 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, Y., Chan, Z., Gao, J., Xing, L., Cao, M., Yu, C., et al. (2016). ABA receptor PYL9 promotes drought resistance and leaf senescence. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1949–1954. doi: 10.1073/pnas.1522840113
- Zsögön, A., Čermák, T., Naves, E. R., Notini, M. M., Edel, K. H., Weinl, S., et al. (2018). *De novo* domestication of wild tomato using genome editing. *Nat. Biotechnol.* 36, 1211–1216. doi: 10.1038/nbt.4272

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 © 2021 Wang, Xun, Wang, Ding, Tian, Hussain, Dong, Li, Cheng, Wang, Lin, Li, Qian, Pang, Feng, Dong, Liu 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.



Identification of Finely Mapped Quantitative Trait Locus and Candidate Gene Mining for the Three-Seeded Pod Trait in Soybean

Candong Li^{1,2,3†}, Hongwei Jiang^{4†}, Yingying Li⁵, Chunyan Liu³, Zhaoming Qi³, Xiaoxia Wu³, Zhanguo Zhang³, Zhenbang Hu³, Rongsheng Zhu³, Tai Guo¹, Zhixin Wang¹, Wei Zheng¹, Zhenyu Zhang¹, Haihong Zhao¹, Nannan Wang^{1,3}, Dapeng Shan³, Dawei Xin^{3*}, Feishi Luan^{2*} and Qingshan Chen^{3*}

¹ Jiamusi Branch Institute, Heilongjiang Academy of Agricultural Sciences, Jiamusi, China, ² College of Life Science, Northeast Agricultural University, Harbin, China, ³ Country College of Agriculture, Northeast Agricultural University, Harbin, China, ⁴ Soybean Research Institute, Jilin Academy of Agricultural Sciences, Changchun, China, ⁵ Key Laboratory of Molecular Epigenetics of MOE, Institute of Genetics and Cytology, Northeast Normal University, Changchun, China

OPEN ACCESS

Edited by:

Madan K. Bhattacharyya,
Iowa State University, United States

Reviewed by:

Ajoy Kumar Roy,
Indian Council of Agricultural
Research (ICAR), India
Yinghui Li,
University of Haifa, Israel

*Correspondence:

Dawei Xin
dwxin@neau.edu.cn
Feishi Luan
luanfeishi@neau.edu.cn
Qingshan Chen
qshchen@126.com

[†] These authors have contributed
equally to this work

Specialty section:

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

Received: 27 May 2021

Accepted: 19 October 2021

Published: 26 November 2021

Citation:

Li C, Jiang H, Li Y, Liu C, Qi Z,
Wu X, Zhang Z, Hu Z, Zhu R, Guo T,
Wang Z, Zheng W, Zhang Z, Zhao H,
Wang N, Shan D, Xin D, Luan F and
Chen Q (2021) Identification of Finely
Mapped Quantitative Trait Locus
and Candidate Gene Mining
for the Three-Seeded Pod Trait
in Soybean.
Front. Plant Sci. 12:715488.
doi: 10.3389/fpls.2021.715488

The three-seeded pod number is an important trait that positively influences soybean yield. Soybean variety with increased three-seeded pod number contributes to the seed number/plant and higher yield. The candidate genes of the three-seeded pod may be the key for improving soybean yield. In this study, identification and validation of candidate genes for three-seeded pod has been carried out. First, a total of 36 quantitative trait locus (QTL) were detected from the investigation of recombinant inbred lines including 147 individuals derived from a cross between Charleston and Dongning 594 cultivars. Five consensus QTLs were integrated. Second, an introgressed line CSSL-182 carrying the target segment for the trait from the donor parent was selected to verify the consensus QTL based on its phenotype. Third, a secondary group was constructed by backcrossing with CSSL-182, and two QTLs were confirmed. There were a total of 162 genes in the two QTLs. The mining of candidate genes resulted in the annotation of eight genes with functions related to pod and seed sets. Finally, haplotype analysis and quantitative reverse transcriptase real-time PCR were carried to verify the candidate genes. Four of these genes had different haplotypes in the resource group, and the differences in the phenotype were highly significant. Moreover, the differences in the expression of the four genes during pod and seed development were also significant. These four genes were probably related to the development process underlying the three-seeded pod in soybean. Herein, we discuss the past and present studies related to the three-seeded pod trait in soybean.

Keywords: soybean, three seeded pods, QTL fine mapping, candidate genes analysis, gene function identification

INTRODUCTION

Traits related to pods and seeds are key factors influencing soybean (*Glycine max* L. Merrill) yield. The soybean yield is determined by evaluating direct yield-related traits at a set planting density, including seed number per plant and 100-seed weight. Seed number per plant, the parameter most related to yield potential (Bárbaro et al., 2006), is closely associated with the number of pods per

plant and with seeds per pod. The seeded-pod trait is a quantitative trait that is highly susceptible to environmental conditions, which leads to ovule abortion and the emergence of various seeded pods on a single plant. This trait also has a certain genetic heritability in different varieties. An increase in the number of three- or four-seeded pods theoretically increases soybean yield. However, a previous study revealed a strong negative correlation between the number of four-seeded pods and 100-seed weight (Carvalho et al., 2017). Another recent study indicated that an increase in the number of four-seeded pods caused decreased seed size, resulting in no yield increase (Li et al., 2018). In soybean production, compared with other varieties, those plants with many three-seeded pods may possess more pods and larger seeds (Zhou et al., 2005; Tavares et al., 2013). The number of three-seeded pods per plant (NThSP) is an important trait for soybean breeding. The “Henong 71” cultivar, which has more three-seeded pods than other varieties (the proportions of two- and three-seeded pods are 10.70 and 82.68%, respectively), produced a record high yield for the Xinjiang Production and Construction Corps of China in 2019. The test yield was 6,712 kg/ha (Guo et al., 2020).

The NThSP is a quantitative trait that is easily affected by the environment. It is difficult to accurately select this trait during breeding. Marker-assisted selection and breeding by design can solve this problem *via* the use of functional molecular markers and genes. Thus, candidate gene mining and subsequent functional verification of genes affecting NThSP are important. Recent advances have been made in quantitative trait locus (QTL) mapping in specific populations, including the gradual improvements of statistical methods and the continual development of genetic maps (Cregan et al., 1999; Song et al., 2004; Choi et al., 2007; Hyten et al., 2010), the gradual enhancement of the technology used for phenotypic analyses of soybean, and increases in the throughput and affordability of genome resequencing. These changes have benefited research aimed at identifying QTL related to NThSP. For example, more than 120 NThSP QTLs are distributed on 20 chromosomes in soybean^{1,2}. These QTLs were detected during investigations of other soybean yield-related traits. To date, NThSP QTLs have not been comprehensively identified or analyzed.

Recently, NThSP QTLs were detected and mapped primarily in recombinant inbred lines (RILs). Several RILs were derived from various hybridizations. First, 15 NThSP QTLs were detected in RILs resulting from the cross between “Jin 23” and “Huibuzhiheidou” (with high three-seeded pod yield). The mean phenotypic variance explained (MPVE) by these QTLs was 1.5–34.70%. Second, an analysis of the RILs derived from the “Zhongdou 29” × “Zhongdou 32” (with high three-seeded pod yield) hybridization detected 13 NThSP QTLs, with an MPVE of 2.54–57.54%. Third, 47 NThSP QTLs were detected in RILs derived from a cross between “Charleston” and “DongNong594” (with a high three-seeded pod yield), with an MPVE of 0.70–46.70%. Finally, two RILs resulting from the “Heihe 36” × “DongnongL 13” (with high three-seeded

pod yield) and “Dongnong L13” (with a high three-seeded pod yield) × “Henong 60” hybridizations were analyzed, resulting in the detection of 20 NThSP QTLs, with an MPVE of 0.71–11.79%. Additionally, four-way RILs were evaluated to identify NThSP QTL (Supplementary Table 1).

Although many NThSP QTLs have been identified, their genomic locations are often undetermined. Moreover, there have been relatively few studies on the candidate gene mining and functional verification of NThSP QTL in soybean. In this study, NThSP QTLs were detected in a RIL population. Meta-QTLs (MQTLs) were subjected to a meta-analysis. The MQTLs were verified using a wild soybean whole-genome introgression line. A secondary group with the target interval was constructed, and the bulked segregant analysis (BSA) resequencing technology was applied for QTL fine mapping and candidate gene mining.

MATERIALS AND METHODS

Plant Materials and Experimental Design

For QTL identification, a RIL comprising 147 individuals was obtained from a cross between an American soybean cultivar (“Charleston” NThSP 20.01) as the female parent (Cooper et al., 1995) and a Chinese cultivar (“DongNong594” NThSP 26.40) as the male parent, which was developed at Northeast Agricultural University, Harbin, Heilongjiang, China. Single seed descent was used to produce F_{2:20}–F_{2:23} generations (Chen et al., 2005). For MQTL validation and secondary group construction, a chromosomal segment substitution line (CSSL) consisting of 220 individuals from the BC₃F₃, BC₃F₄, BC₃F₅, and BC₃F₆ generations was backcrossed with the recurrent parent, “Suinong 14” (developed at Suihua Branch Academy of Heilongjiang Academy of Agricultural Sciences, Suihua, Heilongjiang, China, NThSP 21.87), and the donor parent, “ZYD00006” (wild soybean from the China Germplasm Bank, NThSP 133.20). The linkage group was constructed as described by Xin et al. (2016). Each line in both populations and their parents were grown in Harbin, China (45.75°N, 126.53°E) from 2013 to 2016, according to a randomized complete block design with three replicates. In the experimental plot, the rows were 5 m long, with each row separated by 0.65 m and a plant spacing of 0.05 m.

Data Collection and Quantitative Trait Locus Identification

The traits related to the seed number of pods containing one, two, three, four, and total seeds were recorded as NOSP, NTSP, NThSP, NFSP, and NPPP, respectively. The ratio of the three-seeded pods was determined as the percentage of NThSP to NPPP (Ning et al., 2018). The phenotype data for RILs and their parents were recorded in an earlier study by Li et al., 2018. The average data for five plants per line, randomly selected in each row, were calculated. The QTLs were detected using the composite interval mapping model of the QTL Cartographer (version 2.5) (Wang et al., 2006). The thresholds for detecting QTL were a minimum logarithm of odds score of 2.5 and $P = 0.05$ and were determined using 1,000 permutations at 1-cM intervals. The meta-analysis was performed using tools-meta analysis of the BioMercator

¹<http://www.soybase.org>

²<http://www.ncbi.nlm.nih.gov>

2.1 program. Consensus QTLs were detected according to the optimized model with the lowest Akaike information criterion values. The QTLs were named as previously described (McCouch et al., 1997) (i.e., q + trait name + linkage group number + “-” + QTL number).

Consensus Quantitative Trait Locus Validation and Secondary Group Construction

Targeted lines carrying segments with consensus QTL physical intervals were selected from the 220 CSSLs. The NThSP phenotypic variance between the targeted lines and the recurrent parent was analyzed to validate the consensus QTL. After marker-assisted selection for validating the consensus QTL in the targeted lines, backcrosses with “Suinong 14” (recurrent parent), and two rounds of selfing, we obtained a secondary group comprising 970 BC₄F₂ individuals.

Bulked Segregant Analysis and Fine Mapping

The BSA was used to anchor the candidate intervals (Xia et al., 2010; Song et al., 2016). Two bulks consisting of individuals with extreme phenotypes were selected from the secondary group. The CSSL-182 phenotype of the three-seeded pod trait was 39.22 and that of “Suinong 14” was 21.87. More specifically, 20 plants with more than 34 three-seeded pods (at least 70% of all pods) formed one bulk, whereas the other bulk consisted of 20 plants with fewer than 10 three-seeded pods (less than 30% of all pods). High-quality DNA was extracted from the parental lines and both bulks and used to construct deep-sequencing libraries for BSA. To obtain confident BSA analysis data, specific-locus amplified fragment sequencing (SLAF-seq) was employed, and the sequencing depth was 20 × for the selected soybean materials. Candidate intervals were confirmed using the single nucleotide polymorphism (SNP) index association mapping. The physical positions were determined according to the soybean reference genome (version 1.1) (Schmutz et al., 2010).

Candidate Gene Mining

Candidate genes within the finely mapped QTL regions were predicted as previously described (Li et al., 2018). Interval marker sequences were obtained from the SoyBase database (see text footnote 1). Publicly available resources, including the Kyoto Encyclopedia of Genes and Genomes³ and the Gene Ontology⁴ databases, were used to detect and annotate candidate genes.

Haplotype Analysis

Candidate gene genomic sequences were extracted from the Phytozome database⁵, including the promoter region upstream of the 5′ untranslated region (UTR), coding sequence (CDS), introns, and 3′ UTR. The basic local alignment search tool (BLAST) algorithm was used to compare the candidate

gene genomic sequences with the whole-genome sequence of 92 germplasm resources (10 × sequencing depth), which were grown and phenotypically characterized as RILs and CSSLs. Germplasm resources included the main cultivars in Heilongjiang province, which had abundant phenotypic variation in seeded-pod-related traits. The number of seeded pods was 7.89–38.44, and the ratio of three-seeded pods was 30.99–61.22 (Supplementary Table 2). Haploview 4.2⁶ and Dnasp5.0⁷ were used to analyze the distribution of the main haplotypes in the germplasm resources. The 92 germplasm resources were divided into different classes according to their candidate gene haplotypes. The significance of the phenotypic differences among the haplotype groups of the germplasm resources was determined by ANOVA using the SPSS 20.0 program (IBM Corp., Armonk, NY, United States).

Quantitative Real-Time PCR

The following four materials were selected from the secondary group: lines 215 and 70, which exhibited the high-NThSP-associated trait, and lines 117 and 99, which exhibited the low-NThSP-associated trait (Figure 1B). There were significant phenotypic differences between the two types of materials ($P = 0.05$) (Figure 2C). Flowers, pods, and seeds were collected at the following six stages: beginning flowering (R1), full flowering (R2), beginning pod (R3), full pod (R4), beginning seed (R5), and full seed (R6) (Figure 1A). The collected samples were stored at −80°C for subsequent RNA isolation. All samples were collected as three biological replicates, with each comprising three individual plants. The biological replicates were analyzed in triplicate. The *GmActin4* gene (GenBank ID: AF049106) was used as an internal control. The total RNA was extracted, cDNA was synthesized, and quantitative real-time PCR (qRT-PCR) analysis was performed as previously described (Jiang et al., 2018). The significance of the differences in candidate gene expression levels among the examined stages was determined using SPSS 20.0.

RESULTS

Quantitative Trait Locus Detection and Integration

The NThSP data for the RILs were analyzed in an earlier study (Li et al., 2018). From 2013 to 2016, the average NThSP of “Charleston” was 20.01 and that of “DongNong594” was 26.40. The average NThSP of “Charleston” was 6.39 lower than that of “DongNong594.” The difference in NThSP between the two parents was significant ($P < 0.01$). The NThSP of the RILs varied significantly over the 4-year study period, ranging from 10.48 to 51.19 (Table 1 of Li et al., 2018 and Figure 2A).

A total of 36 NThSP QTLs were detected, which were distributed on GM02, 03, 12, 13, 17, 18, and 19. Of note, 16 QTLs had been published in a previous study (Li et al., 2018), and another 20 QTLs were detected in this study. At least four QTLs at the same location were selected for integration

³ www.genome.jp/kegg/pathway.html

⁴ www.geneontology.org/

⁵ https://phytozome.jgi.doe.gov/pz/portal.html

⁶ https://haploview.software.informer.com

⁷ http://www.ub.edu/dnasp/

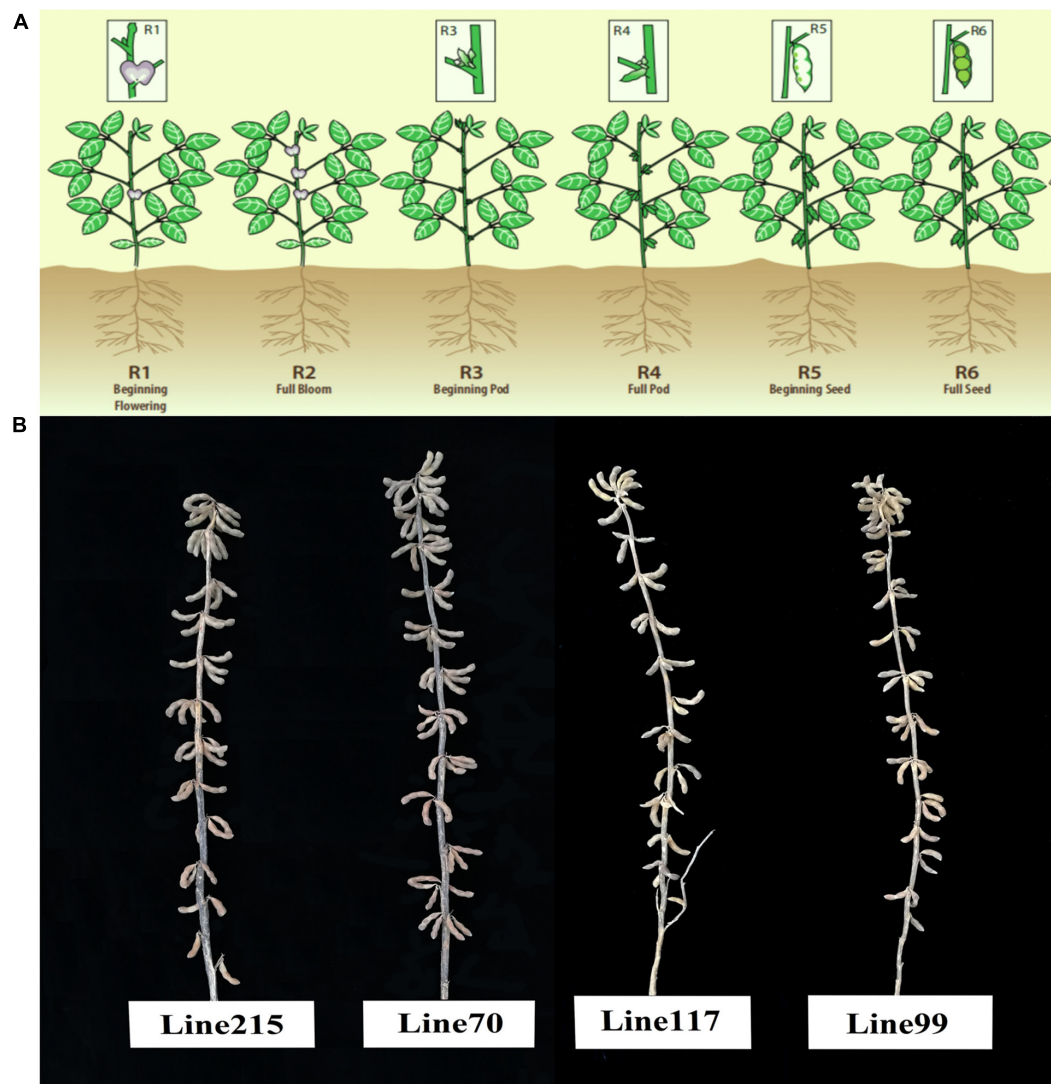


FIGURE 1 | Sampling stages for the quantitative real-time PCR (qRT-PCR) and the individuals with an extreme phenotype regarding the number of three-seeded pods. **(A)** Sampling stages for the qRT-PCR. **(B)** Individuals with an extreme phenotype regarding the number of three-seeded pods. Lines 215 and 70 had a high number of three-seeded pods, whereas lines 117 and 99 had a low number of three-seeded pods.

into a consensus QTL. In this study, five *MQTL* overlapped on GM02, 03, 12, 13, and 17 (**Supplementary Tables 1, 3**), and five NThSP QTL were integrated into *MQTL1*, with a genetic interval of 52.5–54.4 cM and a physical interval of 42.89–46.19 Mb. The additive effects ranged from -7.02 to -2.36 , and the MPVE was 15.74%. Four NThSP QTLs were included in *MQTL2*, with a genetic interval of 1.90–4.80 cM and a physical interval of 2.44–4.63 Mb. The additive effects ranged from -4.35 to -2.25 , and the MPVE was 7.01%. Both *MQTL3* and *MQTL4* comprised four NThSP QTLs, with negative additive effects and an MPVE of 3.01 and 2.97%, respectively. Five NThSP QTLs were included in *MQTL5*, with a genetic interval of 90.1–108.7 cM and a physical interval of 3.49–6.72 Mb. The additive effects ranged from -5.12 to -2.56 , and the MPVE was 10.07% (**Table 1**).

MQTL Validation

In the CSSL population, there were five lines containing donor segments that overlapped with *MQTL1*, CSSL-66, CSSL-75, CSSL-77, CSSL-158, and CSSL-203. Six lines contained donor segments that overlapped with *MQTL2*, CSSL-55, CSSL-123, CSSL-158, CSSL-171, CSSL-182, and CSSL-214. Six lines contained donor segments that overlapped with *MQTL3*, CSSL-46, CSSL-86, CSSL-103, CSSL-151, CSSL-157, and CSSL-182. Five lines contained donor segments that overlapped with *MQTL4*, CSSL-86, CSSL-103, CSSL-105, CSSL-118, and CSSL-182. Five lines contained donor segments that overlapped with *MQTL5*, CSSL-66, CSSL-75, CSSL-103, CSSL-158, and CSSL-182. An analysis of the NThSP data for these CSSL lines and the recurrent parent (Suinong 14) recorded from 2013 to 2016 indicated that the CSSL lines had higher

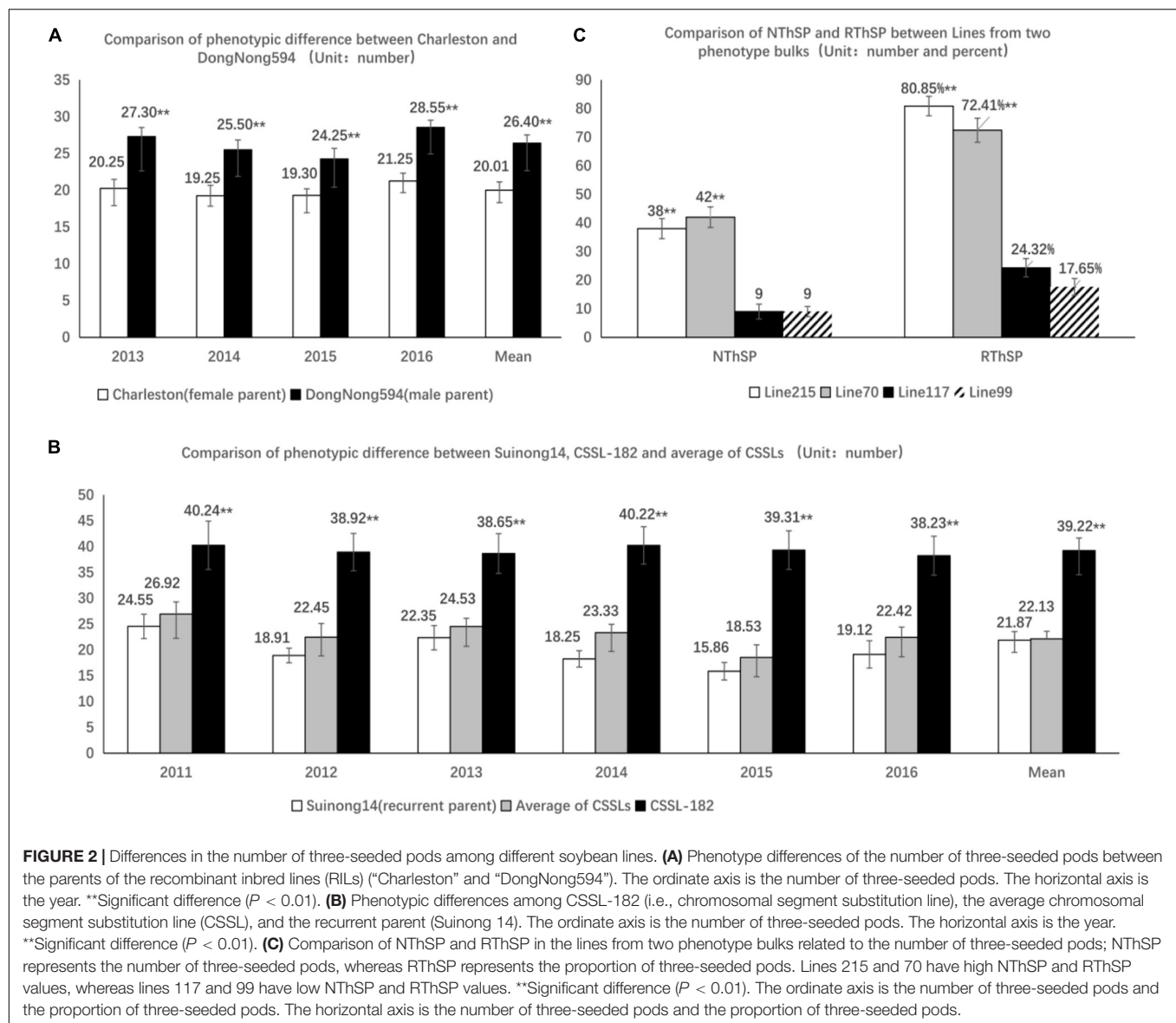


TABLE 1 | The consensus quantitative trait locus (QTL) integration of the number of three-seeded pods.

Consensus QTL	Chromosome	Genetic interval (cM)	Physical interval (Mbp)	QTL numbers	ADD ^a range	Mean PVE ^b (%)
<i>MQTL1</i>	GM02	52.5 ~ 54.4	42.89 ~ 46.19	5	-7.02 ~ -2.36	15.74
<i>MQTL2</i>	GM03	1.90 ~ 4.80	2.44 ~ 4.63	4	-4.35 ~ -2.25	7.01
<i>MQTL3</i>	GM12	26.5 ~ 29.2	10.02 ~ 11.20	4	-2.47 ~ -0.17	3.01
<i>MQTL4</i>	GM13	9.3 ~ 10.3	20.83 ~ 26.06	4	-2.45 ~ -1.42	2.97
<i>MQTL5</i>	GM17	90.1 ~ 108.7	3.49 ~ 6.72	5	-5.12 ~ -2.56	10.07

^aADD was the additive effects contributed by QTL.

^bPVE was the phenotypic variation explained by QTL.

NThSP than “Suinong 14.” The differences were extremely significant over the 4-year study period (**Supplementary Table 4**). This implies that the excellent phenotype of the CSSL lines was the result of the introgression of donor segments that overlapped with the *MQTL*. The *MQTLs* were validated using high generation genome-wide backcrossing

introgression lines based on the phenotype. In particular, CSSL-182 is an excellent material containing donor segments that overlap with four *MQTLs*, namely, *MQTL2*, *MQTL3*, *MQTL4*, and *MQTL5*. The NThSP was significantly higher than in “Suinong 14” as well as the average NThSP for the CSSLs (**Figure 2B**).

Molecular marker analysis of CSSL-182 revealed 8 donor introgression segments, with 319 donor blocks (84.63 Mb) accounting for 8.93% of the whole genome. Moreover, 91.07% of the recurrent parent genome was restored by CSSL-182. The donor introgression segments of CSSL-182 included regions that overlapped with the *MQTL*. On GM03, a 2.19-Mb region overlapped *MQTL2* and the donor block. On GM12, a 1.18-Mb region overlapped *MQTL3* and the donor block. On GM13, a 2.17-Mb region overlapped *MQTL4* and the donor block. On GM17, a 3.23-Mb region overlapped *MQTL5* and the donor block (Table 2). Therefore, CSSL-182 is an ideal material for secondary population construction.

Quantitative Trait Locus Fine Mapping

An F₂ segregating population consisting of 970 individuals was developed by backcrossing CSSL-182 with “Suinong 14.” In this population, the mean NThSP was 34.39 (8.45–61.26), and the mean proportion of pods with three seeds was 44.63% (17.33–82.25%). The absolute skewness and kurtosis values were less than 1.0, suggesting that the segregation of the NThSP phenotype was normally distributed (Table 3).

Two bulks of 20 individuals with extreme phenotypes were selected for BSA to anchor the finely mapped QTL intervals. For the high-NThSP bulk, the NThSP and the proportion of pods with three seeds were 34–57 and 70.27–82.22%, respectively. For the low-NThSP bulk, the NThSP and the proportion of

Pods with three seeds were 6–9 and 17.9–29.17%, respectively (Supplementary Table 5 and Figure 3A1, A2). Two finely mapped QTL intervals were detected on chromosomes GM03 and GM17. For *NThSP03-1*, compared with the reference sequence of CSSL-182, the SNP index of the high-NThSP bulk was 0–1.0 and that of the low-NThSP bulk was 0.1–0.4. The physical interval of *NThSP03-1* was 2.36–3.30 Mb (0.94 Mb), which was contained in the *MQTL2* physical interval 2.44–4.63 Mb (2.19 Mb) and included 42 genes. For *NThSP17-1*, compared with the reference sequence of CSSL-182, the SNP index of the high-NThSP bulk was 0.85–1.0, and the low-NThSP bulk was 0.15–0.4. The physical interval of *NThSP17-1* was 4.34–5.68 Mb (1.34 Mb), which was contained in the *MQTL5* physical interval 3.49–6.72 Mb (3.23 Mb) and included 120 genes. Two finely mapped QTL intervals included 162 genes. On chromosome GM03, the candidate genes could be divided into three sections, namely, pollen germination, mRNA modification, and ATP-dependent helicase activity. The genes found on chromosome GM17 could then be divided into different sections depending on the gene function annotation, such as cellulose synthesis, metabolic processes, or other biochemical pathways (Figure 3B1, B2 and Supplementary Table 6).

Candidate Gene Prediction

Previous research has shown that ovule, pollen, stamen, embryo, and flower development considerably influence pod and seed sets

TABLE 2 | The information of CSSL-182 and the overlap regions with the *MQTL* of the number of the three-seeded pod.

Code	Block number	Segments length of Block (Mbp)	Introgressed ratio (%)	Interval of molecular marker	Consensus QTL	Physical intervals of <i>MQTL</i> (Mbp)	Physical intervals of block (Mbp)	Intersection intervals (Mbp)	Chromosome
1	25	6.22	0.66	Block1481 ~ Block1565	<i>MQTL2</i>	2.44 ~ 4.63	0.24 ~ 6.46	2.19	GM03
2	71	10.69	1.13	Block6225 ~ Block6245	<i>MQTL3</i>	10.02 ~ 11.20	7.21 ~ 17.90	1.18	GM12
3	35	5.84	0.62	Block6666 ~ Block6684	<i>MQTL4</i>	20.83 ~ 26.06	23.89 ~ 29.73	2.17	GM13
4	31	3.38	0.36	Block8733 ~ Block8771	<i>MQTL5</i>	3.49 ~ 6.72	3.43 ~ 6.81	3.23	GM17
Else	157	58.50	6.17						GM02, 06, 19, 20
Total	319	84.63	8.93	4	4	11.83	26.13	8.77	8

TABLE 3 | Statistical analysis of traits related to the number of the three-seeded pod in three populations.

Population	Trait	Min.	Max.	Mean ± SD	Range	Skewness	Kurtosis
RILs	NThSP	10.48	51.19	26.77 ± 7.87	10.48 ~ 51.19	0.51	0.42
	RThSP	27.26	65.22	47.25 ± 7.03	27.26 ~ 65.22	-0.22	0.09
	NPPP	26.50	96.51	56.34 ± 13.13	26.50 ~ 96.51	0.55	0.53
CSSLs	NThSP	12.32	49.35	22.13 ± 3.64	12.32 ~ 49.35	0.99	1.06
	RThSP	19.24	75.21	45.22 ± 5.32	19.24 ~ 75.21	1.22	0.78
	NPPP	24.17	98.32	68.43 ± 8.21	24.17 ~ 98.32	1.78	1.39
Secondary group	NThSP	6.00	61.00	34.39 ± 2.34	6.00 ~ 61.00	0.63	0.03
	RThSP	17.33	82.25	44.63 ± 3.55	17.33 ~ 82.25	0.66	0.21
	NPPP	23.00	86.00	54.66 ± 6.35	23.00 ~ 86.00	-0.28	0.49
Germplasm resources	NThSP	7.89	38.44	21.65 ± 6.61	7.89 ~ 38.44	0.51	-0.23
	RThSP	30.99	61.22	46.22 ± 7.35	30.99 ~ 61.22	-0.06	-0.69
	NPPP	24.56	78.33	46.45 ± 10.40	24.56 ~ 78.33	0.24	0.08

NThSP indicates the trait of the number of the three-seeded pod. *RThSP* indicates the ratio of the number of the three-seeded pod. *NPPP* indicates the trait of the number of pods per plant. *RILs* indicate recombinant inbred lines. *CSSLs* indicate chromosomal segment substitution lines. *Min* and *Max* indicate the minimum and maximum values, respectively.

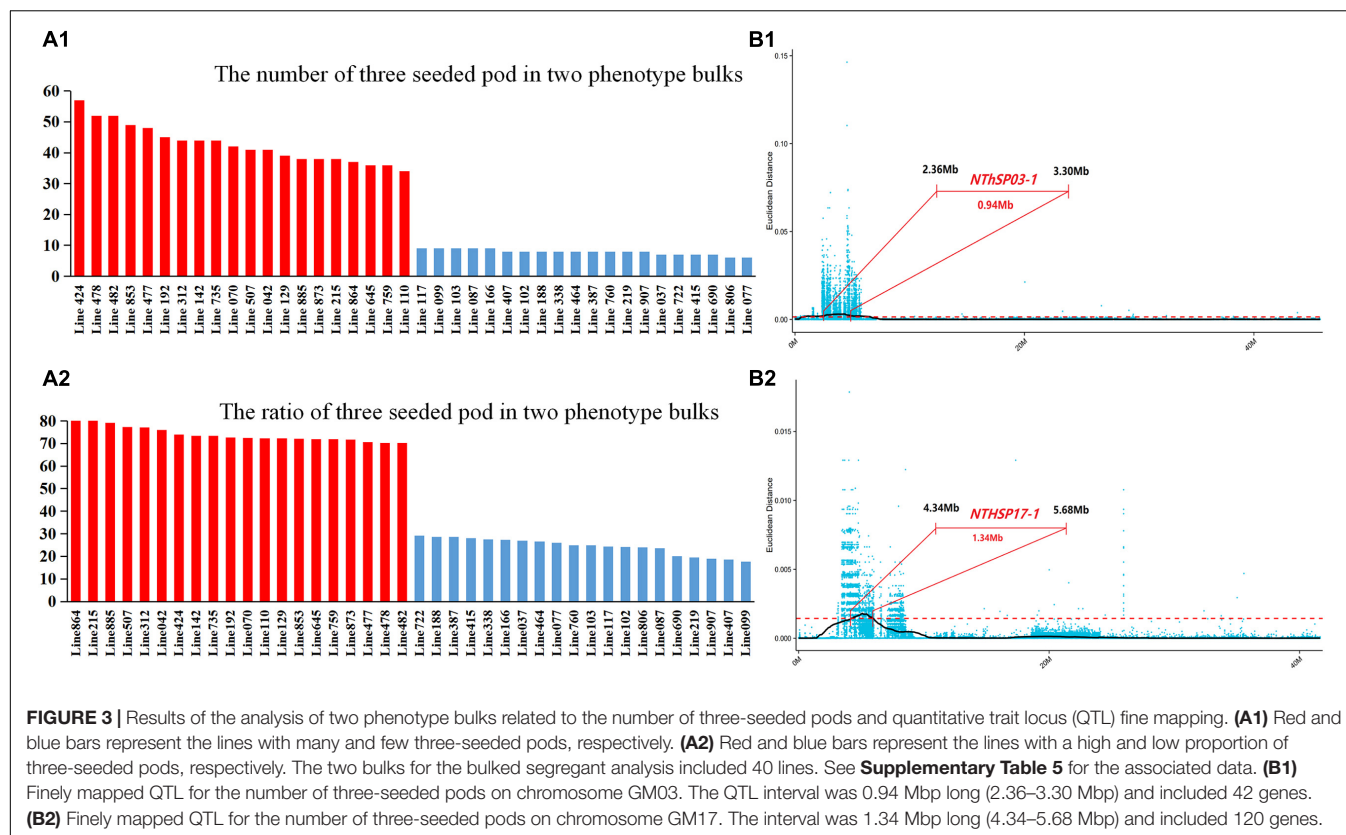


FIGURE 3 | Results of the analysis of two phenotype bulks related to the number of three-seeded pods and quantitative trait locus (QTL) fine mapping. **(A1)** Red and blue bars represent the lines with many and few three-seeded pods, respectively. **(A2)** Red and blue bars represent the lines with a high and low proportion of three-seeded pods, respectively. The two bulks for the bulked segregant analysis included 40 lines. See **Supplementary Table 5** for the associated data. **(B1)** Finely mapped QTL for the number of three-seeded pods on chromosome GM03. The QTL interval was 0.94 Mbp long (2.36–3.30 Mbp) and included 42 genes. **(B2)** Finely mapped QTL for the number of three-seeded pods on chromosome GM17. The interval was 1.34 Mbp long (4.34–5.68 Mbp) and included 120 genes.

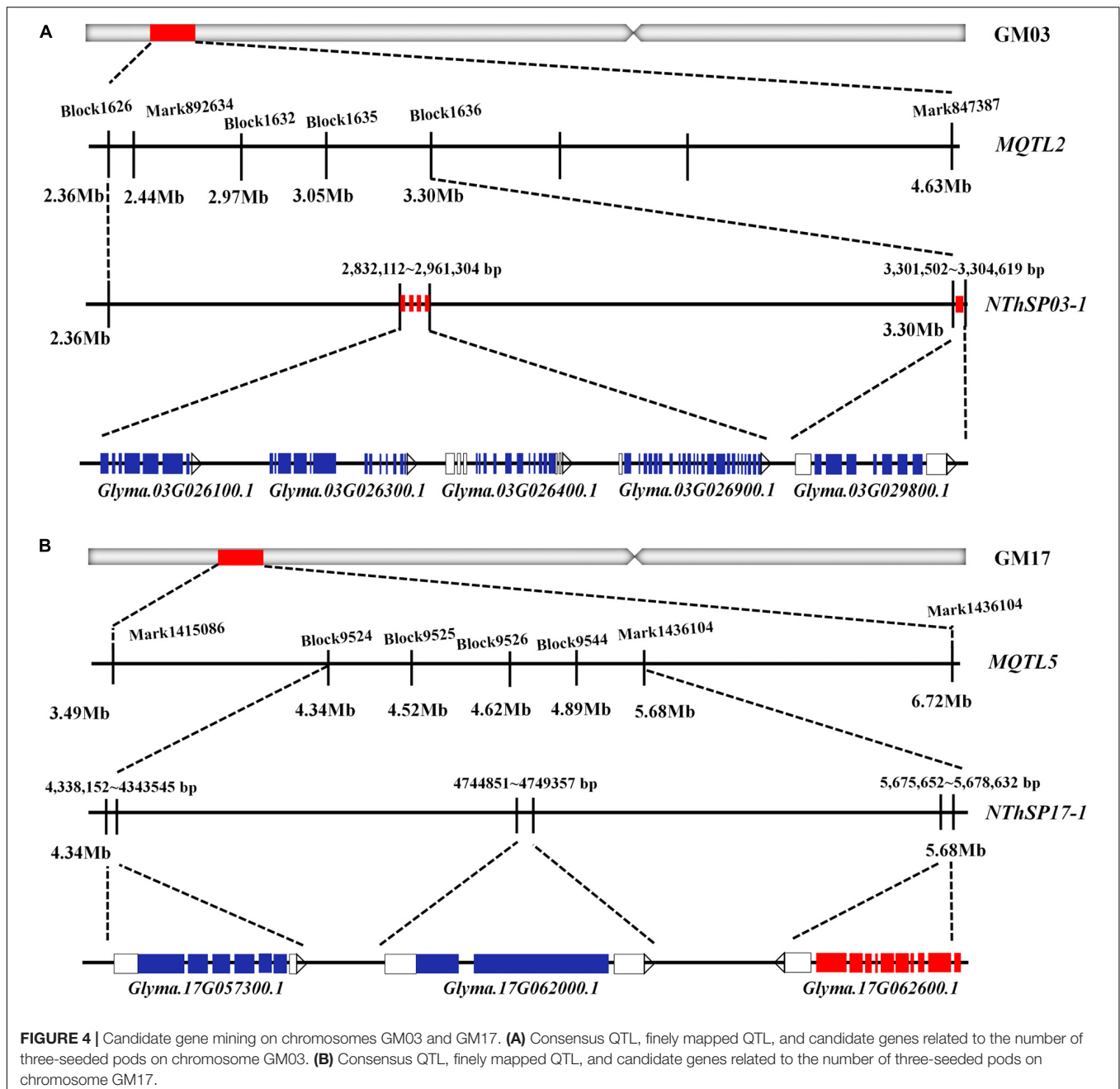
in the soybean (Mukhtar and Coyne, 1981; Tischner et al., 2003; Mena-Alí and Rocha, 2005; Endo and Ohashi, 2009; Kurdyukov et al., 2014; Jiang et al., 2015; Qi et al., 2015). Gene functions related to these biological processes may affect NThSP. Based on this, 8 candidate genes were selected from 162 genes in the target intervals. Homologous gene analysis was performed in the soybean and *Arabidopsis* genomes. For eight genes, five candidate genes were detected in the *NThSP03-1* interval (i.e., *Glyma.03G026100.1*, *Glyma.03G026300.1*, *Glyma.03G026400.1*, *Glyma.03G026900.1*, and *Glyma.03G029800.1*). The homologous genes in the soybean genome were *Glyma.01G140600.1* and *Glyma.17G138200.1* (Zhao et al., 2017; Chen et al., 2018), whereas the homologous genes in the *Arabidopsis* genome were *AT1G71820* and *AT2G45190* (Nolan et al., 2009; Lenser et al., 2016; Pathak et al., 2016). Based on previous studies, functional characterization of these genes indicated that they are involved in pollen germination; pollen tube growth; meristem growth; ovule, embryos, flower, and fruit development; and cell division. Another three candidate genes were included in the *NThSP17-1* interval (*Glyma.17G057300.1*, *Glyma.17G062000.1*, and *Glyma.17G062600.1*). The homologous genes in the soybean genome are *Glyma.13G101900.1*, *Glyma.13G097600.1*, and *Glyma.13G096900.1* (Wang et al., 2015), whereas the homologous genes in the *Arabidopsis* genome are *AT1G73590*, *AT5G57360*, and *AT5G57390* (Bai and Demason, 2008; Kwak et al., 2008; Tsuwamoto et al., 2010; Xue et al., 2012; Hu et al., 2014; Liew et al., 2014; Radoeva and Weijers, 2014; Seefried et al., 2014). Based on previous studies, functional characterization

of these genes revealed that they are related to embryo, flower, stamen, embryo development, and seed germination. The information for each gene and its homologous gene from the literature are listed in **Table 4**. Therefore, these eight genes may be important candidate genes for NThSP (**Table 4** and **Figure 4**).

Candidate Gene Analysis

Haplotype and qRT-PCR analyses were performed to assess the candidate genes. Of the eight predicted candidate genes, four had different haplotypes in the germplasm resources. The NThSP of the highest line was 38.44 and that of the lowest line was 7.89. The mean NThSP was 21.65 in the germplasm resources. The traits related to the three-seeded pod were approximately normally distributed in the germplasm resources (**Table 3**). The SNPs of the haplotypes for each candidate gene are listed in **Table 5**. The NThSP phenotypic differences among the different haplotypes were significant, and four genes were differentially expressed during pod and seed development. The expression levels of the candidate genes reached significant levels during the six developmental periods. The four candidate genes were *Glyma.03G029800.1*, *Glyma.17G057300.1*, *Glyma.17G062000.1*, and *Glyma.17G062600.1* (**Tables 5, 6**).

For the haplotypes of the four candidate genes, the phenotypes of the high-NThSP lines were significantly higher than the low-NThSP lines in the germplasm resources. In *Glyma.03G029800.1*, the NThSP and proportion of pods with three seeds were significantly higher for Hap-1 (22.96 and 43%) and Hap-17 (23.24 and 42%) than for Hap-7 (17.52 and 33%). The SNPs



were located upstream of the promoter at -1,568, -564, and -479 bp. In *Glyma.17G057300.1*, the NThSP and proportion of pods with three seeds were significantly higher for Hap-3 (25.58 and 42%) than for Hap-2 (20.01 and 36%) and Hap-6 (21.26 and 38%). The SNPs were located upstream of the promoter at -1,701 and -1,526 bp and in the CDS region of 1,117 bp. In *Glyma.17G062000.1*, the NThSP and proportion of pods with three seeds were significantly higher for Hap-2 (26.37 and 49%) than for Hap-1 (21.86 and 42%) and Hap-3 (21.38 and 43%). The SNPs were located upstream of the promoter at -2,175 bp and -455 bp and in the CDS region of 2,501 bp and 3,128 bp. In *Glyma.17G062600.1*, the NThSP and proportion of pods with

three seeds were significantly higher for Hap-1 (25.08 and 51%) than for Hap-3 (20.47 and 45%). The SNPs were located upstream of the promoter at -2,938 bp, -2,758 bp, and 874 bp and in the CDS region of 1,963 bp and 2,890 bp. The phenotypic differences between the haplotypes for each candidate gene may be the result of these SNP differences (Table 5).

The qRT-PCR analysis indicated that *Glyma.03G029800.1* was more highly expressed in lines 215 and 70 (i.e., the high-NThSP lines) than in lines 117 and 99 (i.e., the low-NThSP lines) or Suinong 14 (recurrent parent) in the R1, R2, R3, R4, and R5 stages. *Glyma.17G057300.1* was more highly expressed in lines 215 and 70 than in lines 117 and 99 or Suinong 14 in the R1, R2,

TABLE 4 | Candidate genes and their homologous genes in soybean and *Arabidopsis* genome.

Code	Candidate genes	Soybean homologous	References	Arabidopsis homologous	References	Common annotation information of biological process
1	<i>Glyma.03G026100.1</i>	<i>Glyma.01G140600.1</i>	Chen et al., 2018	<i>AT1G71820</i>	Kim et al., 2009; Safavian et al., 2015	Pollen germination, pollen tube growth
2	<i>Glyma.03G026300.1</i>	<i>Glyma.01G140600.1</i>	Chen et al., 2018	<i>AT1G71820</i>	Kim et al., 2009; Safavian et al., 2015	Pollen germination, pollen tube growth
3	<i>Glyma.03G026400.1</i>	<i>Glyma.01G140600.1</i>	Chen et al., 2018	<i>AT1G71820</i>	Kim et al., 2009; Safavian et al., 2015	Pollen germination, pollen tube growth
4	<i>Glyma.03G026900.1</i>	<i>Glyma.01G140600.1</i>	Chen et al., 2018	<i>AT1G71820</i>	Kim et al., 2009; Safavian et al., 2015	Pollen germination, pollen tube growth
5	<i>Glyma.03G029800.1</i>	<i>Glyma.17G138200.1</i>	Zhao et al., 2017	<i>AT2G45190</i>	Lenser et al., 2016; Pathak et al., 2016	Ovule, embryo, flower, fruit development, meristem growth
6	<i>Glyma.17G057300.1</i>	<i>Glyma.13G101900.1</i>	Wang et al., 2015; Stasko, 2018	<i>AT1G73590</i>	Fang, 2008; Hu et al., 2014; Seefried et al., 2014	Embryo development, flower and stamen development
7	<i>Glyma.17G062000.1</i>	<i>Glyma.13G097600.1</i>	-	<i>AT5G57360</i>	Kwak et al., 2008; Xue et al., 2012; Liew et al., 2014	Flower development
8	<i>Glyma.17G062600.1</i>	<i>Glyma.13G096900.1</i>	-	<i>AT5G57390</i>	Tsuwamoto et al., 2010; Radoeva and Weijers, 2014	Positive regulation of embryonic development, 'seed germination

and R4 stages. The expression of *Glyma.17G062000.1* was lower in lines 215 and 70 than in lines 117 and 99 or Suinong 14 in the R1, R2, R3, and R4 stages. *Glyma.17G062600.1* was more highly expressed in line 215 than in Suinong 14 in the R1, R2, R3, and R4 stages. This gene was more highly expressed in line 70 than in Suinong 14 in the R3 and R4 stages. Moreover, the expression level of *Glyma.17G062600.1* was higher in lines 215 and 70 than in lines 117 and 99 in the R3, R4, and R6 stages (Table 6).

DISCUSSION

The results of this study corroborate those reported in previous studies. For example, the *MQTL1* region on GM02 (42.89–46.19 Mbp) included five QTLs, all of which had negative effects, suggesting that they were stable QTLs throughout the study period. An earlier investigation detected a QTL (*qPN-D1b-2*) using a Satt546 simple sequence repeat marker (physical position: 43,775,564–43,775,623 bp) associated with seed set and seed yield (Ning et al., 2018). Another QTL (*qPN-D1b-3*) detected using Sat_183 (physical position: 44,317,044–44,317,314 bp) and Sat_069 (physical position: 46,353,731–46,353,789 bp) was associated with seed set and seed weight (Ning et al., 2018). A QTL (*QNTPD1b-1*) associated with Sat_135 (physical position: 40,366,215–40,366,272 bp) was reported to be related to the two-seeded pod trait (Asakura et al., 2012; Yang et al., 2013b). Other studies have shown that QTL *qSN-1*, which is associated with Satt350 (physical position: 40,366,215–40,366,272 bp), is also related to the two-seeded pod trait (Asakura et al., 2012; Yang et al., 2013b). Additionally, the QTL *qSN-1*, which is close to the *MQTL1* region on GM02, is associated with Satt189 and Satt350 and is related to early flowering and seed number (Li et al., 2010). Thus, *MQTL1* has been identified by several studies as a genomic locus influencing seed sets. This is an important consensus QTL interval. The *MQTL2* region on

GM03 (2.44–4.63 Mbp) includes four QTLs. A previous study revealed that a QTL (*qfn-Chr3*) associated with Satt009 (physical position: 3,910,260–3,910,307 bp) is within the *MQTL2* interval and is related to flower number (Zhang et al., 2010). No loci consistent with *MQTL3* have been reported; thus, this is a new region related to NThSP. The *MQTL4* region on GM13 (20.83–26.06 Mbp) contains a QTL related to the number of ovules per pod; this QTL is associated with Sat_133 (physical position: 23,462,623–23,462,676 bp) (Tischner et al., 2003). Another QTL (*qPN-D2-2*), which is associated with Sat_284 (physical position: 6,551,297–6,551,348 bp) in the *MQTL5* interval on GM17 (3.49–6.72 Mbp), is related to the seed set (Ning et al., 2018). The QTLs identified in earlier studies are associated with flower number, number of ovules per pod, pod, and seed set. These loci are consistent with the consensus QTL identified in this study, which may facilitate the prediction of candidate genes related to NThSP.

Soybean seed yield is positively associated with the number of flowers as well as successful pod and seed set (Egli, 2005; Pushpavalli et al., 2015). The number of seeds per pod is related to the number of ovules that develop in each pod. There are usually five ovules in one flower, but the number of ovules that are aborted ultimately determines the number of seeds in a pod (Shibles et al., 1975; Tischner et al., 2003). Moreover, pollen tube growth, stamen development, and flower development are also important factors (Mukhtar and Coyne, 1981; Tischner et al., 2003; Mena-Alí and Rocha, 2005; Endo and Ohashi, 2009; Kurdyukov et al., 2014; Jiang et al., 2015; Qi et al., 2015; Smitha Ninan et al., 2017). Accordingly, candidate genes with annotated functions related to pollen, stamen, ovule, embryo, and flower development as well as pollen tube growth are likely important for pod and seed set.

In this study, eight genes were annotated with functions associated with pod and seed sets, and homologous genes in the soybean and *Arabidopsis* genomes were functionally

TABLE 5 | Haplotype analysis of the candidate genes of the number of the three-seeded pod.

(A)													
Candidate genes	Haplotype code	Phenotypic types	Ratio of haplotype type	Mean value of NThSP \pm SD	Mean value of RThSP \pm SD	Upstream promoter –1568bp	Upstream promoter –564bp	Upstream promoter –479bp	Upstream promoter –2175bp	Upstream promoter –455bp	CDS 2501bp	CDS 3128bp	
Glyma.03G029800.1	Hap_1	High	25.0%	22.96 \pm 5.34 a	0.43 \pm 0.04 a	A	A	A					
	Hap_17	High	31.5%	23.24 \pm 6.45 a	0.42 \pm 0.05 a	A	T	A					
	Hap_7	Low	27.2%	17.52 \pm 3.68 c	0.33 \pm 0.08 c	G	A	G					
	Else		16.3%	20.15 \pm 2.85 b	0.38 \pm 0.12 b								
Glyma.17G062000.1	Hap_2	High	29.4%	26.37 \pm 5.77 a	0.49 \pm 0.07 a				G	C	A	A	
	Hap_1	Low	41.3%	21.86 \pm 5.39 c	0.42 \pm 0.08 c				A	G	T	T	
	Hap_3	Low	23.9%	21.38 \pm 8.21 c	0.43 \pm 0.02 b				G	G	T	T	
	Else		5.4%	23.89 \pm 4.52 b	0.45 \pm 0.09 b								
(B)													
Candidate genes	Haplotype code	Phenotypic types	Ratio of haplotype type	Mean value of NThSP \pm SD	Mean value of RThSP \pm SD	Upstream promoter –1701bp	Upstream promoter –1526bp	CDS 1117bp	Upstream promoter –2938bp	Upstream promoter –2758bp	Upstream promoter –874bp	CDS 1963bp	CDS 2890bp
Glyma.17G057300.1	Hap_3	High	23.9%	25.58 \pm 5.54 a	0.42 \pm 0.03 a	A	T	A					
	Hap_2	Low	42.4%	20.01 \pm 3.25 bc	0.36 \pm 0.02 c	C	C	G					
	Hap_6	Low	26.1%	21.26 \pm 6.34 b	0.38 \pm 0.03 c	C	T	G					
	Else		7.6%	22.85 \pm 6.41 b	0.40 \pm 0.09 b								
Glyma.17G062600.1	Hap_1	High	28.3%	25.08 \pm 6.34 a	0.51 \pm 0.18 a				C	A	A	A	C
	Hap_3	Low	56.5%	20.47 \pm 3.21 c	0.45 \pm 0.21 c				G	G	T	T	G
	Else		15.2%	22.58 \pm 3.54 b	0.48 \pm 0.15 b								

NThSP represents the number of the three-seeded pod, whereas RThSP represents the proportion of the three-seeded pod. The different lowercase letters in columns of each candidate gene indicate the significant difference level ($P < 0.05$).

SNPs are indicated by colored letters.

TABLE 6 | Expression analysis of the candidate genes of the number of the three-seeded pod.

Candidate genes	Extremely lines	Phenotypic types	Period and expression indicators					
			R1	R2	R3	R4	R5	R6
Glyma.03G029800.1	Suinong14	CK	5.60 ab	2.70 b	6.77 b	3.39 bc	4.91 ab	5.61 a
	Line215	High	7.60 a	4.65 a	8.55 a	5.93 a	5.32 a	4.36 b
	Line70	High	6.42 a	5.54 a	9.16 a	5.00 a	6.90 a	5.25 a
	Line117	Low	3.49 b	1.93 c	4.96 bc	2.35 c	3.97 c	2.70 c
	Line99	Low	3.93 b	1.18 c	2.27 c	4.61 b	4.10 b	3.29 bc
Glyma.17G057300.1	Suinong14	CK	3.23 b	4.33 a	5.25 a	4.77 ab	4.67 ab	3.92 ab
	Line215	High	5.08 a	5.32 a	6.42 a	5.06 a	5.25 a	5.15 a
	Line70	High	4.04 a	4.60 a	5.42 ab	5.61 a	5.05 a	4.71 a
	Line117	Low	2.68 b	2.32 b	4.42 b	3.06 b	4.25 ab	4.15 a
	Line99	Low	2.04 b	3.02 b	3.73 ab	2.11 c	3.22 b	3.71 b
Glyma.17G062000.1	Suinong14	CK	2.61 ab	2.37 b	4.46 ab	2.34 ab	2.92 ab	2.42 b
	Line215	High	1.44 b	1.18 c	2.60 b	1.77 b	2.43 b	2.30 b
	Line70	High	1.21 b	1.42 c	3.31 b	1.30 b	2.75 b	2.72 b
	Line117	Low	3.91 a	3.68 ab	5.82 a	3.64 a	3.08 a	3.53 a
	Line99	Low	2.91 ab	4.22 a	4.93 a	3.41 a	3.57 a	3.03 ab
Glyma.17G062600.1	Suinong14	CK	4.53 b	3.43 b	5.18 b	3.13 b	4.14 ab	5.61 ab
	Line215	High	6.92 a	5.67 a	7.42 a	6.02 a	5.11 a	6.12 a
	Line70	High	3.52 c	4.32 ab	6.17 a	5.60 a	4.97 a	7.02 a
	Line117	Low	2.55 cd	3.22 b	2.43 c	2.51 c	5.23 a	3.09 c
	Line99	Low	3.04 c	3.95 b	2.90 c	1.96 c	3.55 b	4.30 b

The different lowercase letters in columns of each candidate gene indicate the significant difference level ($P < 0.05$).

annotated. In the *NThSP03-1* interval, four candidate genes (i.e., *Glyma.03G026100.1*, *Glyma.03G026300.1*, *Glyma.03G026400.1*, and *Glyma.03G026900.1*) were identified as homologs in the soybean and *Arabidopsis* genomes. Based on circular RNA differential expression analyses, these genes were found to affect pollen germination and pollen tube growth (Nolan et al., 2009; Safavian et al., 2015; Chen et al., 2018). However, excellent haplotypes for these genes were not detected in the germplasm resources included in this study, implying that these genes are expressed ubiquitously in soybean and are unrelated to seed sets.

The *Glyma.03G029800.1* gene was identified as *GmYABBY5*, which helps maintain meristems during the early floral developmental stages (Zhao et al., 2017). The *Arabidopsis* homolog of this gene is *AT2G45190*, which mediates biological processes related to ovule, embryo, and flower development (Lenser et al., 2016; Pathak et al., 2016). In tomatoes, the *YABBY*-like transcription factor was also found to play a pivotal role in fruit size regulation (Cong et al., 2008). Similar functions of *YABBY5* were also identified in the gynoecium of rice and sugar-apple (Yang and Hwa, 2008). Additionally, this gene underlies the development of ovules (Lora et al., 2011; Cucinotta et al., 2020), supporting that it might also have similar functions to *NThSP*. In this study, three excellent *Glyma.03G029800.1* haplotypes were detected in the examined germplasm resources. The *NThSP* data differed significantly among these three haplotypes, indicating that this gene influences the *NThSP* phenotype. The CAAT-box is a common *cis*-acting promoter element that is recognized by *trans*-acting factors that increase promoter activity. In the *Glyma.03G029800.1* promoter region, a base change from A to G

at −1,568 and −479 bp adversely affects the CAAT-box promoter element, which may explain the observed decrease in *NThSP*. Additionally, the detected base changes in the promoter regions of *Glyma.17G057300.1* and *Glyma.17G062600.1* were also within the CAAT-box element. The different haplotypes of these two genes were associated with similar phenotypic changes. The pod and seed sets may be closely related to increased gene expression resulting from a functional CAAT-box element. The qRT-PCR analysis provided additional information regarding the potential gene functions. High *Glyma.03G029800.1*, *Glyma.17G057300.1*, and *Glyma.17G062600.1* expression levels may promote pod and seed sets. There were significant differences in the expression levels of these genes between the lines with a high *NThSP* and those with a low *NThSP*.

The *Glyma.17G057300.1* gene was identified as *GmPIN2b*. A previous study confirmed that upregulated *GmPIN2b* expression can induce soybean flower growth and grain development (Wang et al., 2015). The *AT1G73590* gene is the *Arabidopsis* homolog of *PIN1* and *Glyma.17G057300.1*. This gene mediates *Arabidopsis*, tomato, and soybean flower and fruit development. Earlier research confirmed that *AT1G73590* expression affects auxin transport in embryos (Bai and Demason, 2008; Hu et al., 2014; Seefried et al., 2014). Accordingly, *Glyma.17G057300.1* may be another candidate gene related to *NThSP*. In this study, different *Glyma.17G057300.1* haplotypes were detected in the analyzed germplasm resources, with very significant differences in the phenotypes associated with the haplotypes. In the CDS of this gene, the A base at 1,117 bp resulted in an aspartic acid in the encoded protein as well as

a high NThSP, whereas a G at this position resulted in glycine and a low NThSP. The base change in *Glyma.17G062600.1* CDS (2,890 bp) similarly affected NThSP. More specifically, a C at this position resulted in alanine and a high NThSP, whereas a G at this position resulted in glycine and a low NThSP. The qRT-PCR data may be useful for functionally characterizing *Glyma.17G057300.1* and *Glyma.17G062600.1*.

In this study, a RIL was used to identify QTL potentially related to NThSP. A CSSL was used to verify the candidate QTL intervals. A secondary group was used to map the fine QTL. Another germplasm resource population was used to check the candidate genes by haplotype identification. Comparing four populations, the range of NThSP and three-seeded trait values in the RILs, CSSLs, and secondary group was greater than in the germplasm resources. This indicated evident trait separation in the targeted populations. However, the haplotype analysis of candidate genes in the germplasm resources was more objective. The haplotypes and NThSP phenotype could be better validated in the germplasm resources. In this study, most of the higher NThSP phenotype lines, including the highest one, had the same haplotype. For example, there were 27 lines, accounting for 29.35% of the germplasm resources, with a higher NThSP phenotype that had the same haplotype (Table 5). This suggested that the candidate gene haplotype was probably related to the NThSP phenotype. However, it is possible that there were few higher NThPP lines that did not share the same haplotype with it. These lines with different haplotypes and higher NThPP phenotypes belonged to the else Haps, which could be attributed to other genes. It is likely that other NThSP genes exist in the study population. Additionally, CSSL-182, which included the genomic segments imported from the donor, had a better phenotype than the recurrent parent. This was conducive to the segregation and recombination of the target intervals in the secondary group and the fine mapping of NThSP QTL. Candidate genes were selected strictly according to functional annotations. Thus, there may be other candidate genes related to NThSP in the QTL intervals. The selected candidate genes were annotated with functions related to pollen tube growth, as well as the development of pollen grains, stamens, ovules, embryos, and flowers. An analysis of the candidate genes revealed that they are closely related to pods and seed sets. Future studies should more precisely characterize the functions of the identified candidate genes in transgenic soybean plants. In addition, there were different haplotypes in the germplasm resources for the candidate genes. Several SNPs existed in the different haplotypes. These

SNPs could be developed into KASP (competitive allele-specific polymerase chain reaction) markers, and the KASP markers could be evaluated in the segregated populations to confirm their function for the selection of high-NThSP materials in molecular marker-assisted 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

QC and FL: conceptualization. ZQ and HJ: methodology. DX, HJ, and XW: validation. CDL, TG, ZW, CYL, and NW: formal analysis. WZ, HZ, ZYZ, RZ, and DS: data curation. CDL: writing-original draft. DX, HJ, and QC: writing-review and editing. ZH: project administration. All authors have read and agreed to the published version of the manuscript.

FUNDING

This study was financially supported by the National Natural Science Foundation of China (31801389), the Natural Science Foundation of Heilongjiang Province (YQ2019C021), and the Postdoctoral Foundation of Heilongjiang Province (LBH-Z16185).

ACKNOWLEDGMENTS

We would like to thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.715488/full#supplementary-material>

REFERENCES

- Asakura, T., Tamura, T., Terauchi, K., Narikawa, T., Yagasaki, K., Ishimaru, Y., et al. (2012). Global Gene Expression Profiles in Developing Soybean Seeds. *Plant Physiol. Biochem.* 52, 147–153. doi: 10.1016/j.plaphy.2011.12.007
- Bárbaro, I. M., Mauro, A., Unédatrevisoli, S. H., Arriel, N., and Costa, M. M. (2006). Path Analysis and Expected Response in Indirect Selection for Grain Yield in Soybean. *Crop Breed. Appl. Biot.* 6, 151–159. doi: 10.12702/1984-7033.v06n02a06
- Carvalho, I. R., Nardino, M., Demari, G. H., Szareski, V. J., Follmann, D. N., Pelegri, A. J., et al. (2017). Relations Among Phenotypic Traits of Soybean Pods and Growth Habit. *Afr. J. Agri. Res.* 6, 450–458. doi: 10.5897/AJAR2016.11660
- Chen, L. F., Ding, X. L., Zhang, H., He, T. T., Li, Y. W., Wang, T. L., et al. (2018). Comparative Analysis of Circular RNAs Between Soybean Cytoplasmic Male-sterile Line NJCMS1A and Its Maintainer NJCMS1B by High-throughput Sequencing. *BMC Genomics* 19:663. doi: 10.1186/s12864-018-5054-6
- Chen, Q. S., Zhang, Z. C., Liu, C. Y., Wang, W. Q., and Li, W. B. (2005). Construction and Analysis of Soybean Genetic Map Using Recombinant Inbred Line of Charleston × Dongnong 594. *Sci. Agri. Sin.* 12, 13–18.

- Choi, I. Y., Hyten, D. L., Matukumalli, L. K., Song, Q., Chaky, J. M., Quigley, C. V., et al. (2007). A Soybean Transcript Map: gene Distribution, Haplotype and Single-nucleotide Polymorphism Analysis. *Genet* 176, 685–696. doi: 10.1534/genetics.107.070821
- Cong, B., Ba Rrero, L. S., and Tanksley, S. D. (2008). Regulatory Change in Yabby-Like Transcription Factor Led to Evolution of Extreme Fruit Size During Tomato Domestication. *Nat. Genet.* 40, 800–804. doi: 10.1038/ng.144
- Cooper, R. L., Martin, R. J., and Martin, S. K. (1995). Registration of ‘Charleston’ Soybean. *Crop Sci.* 35:592. doi: 10.2135/cropsci1995.0011183X003500020060x
- Cregan, P. B., Jarvik, T., Bush, A. L., Shoemaker, R. C., Lark, K. G., Kahler, A. L., et al. (1999). An Integrated Genetic Linkage Map of the Soybean. *Crop Sci.* 39, 1464–1490. doi: 10.2135/cropsci1999.3951464x
- Cucinotta, M., Di Marzo, M., Guazzotti, A., de Folter, S., Kater, M. M., and Colombo, L. (2020). Gynoecium Size and Ovule Number are Interconnected Traits That Impact Seed Yield. *J. Exp. Bot.* 71, 2479–2489. doi: 10.1093/jxb/eraa050
- Du, J., Li, W. X., Dong, Q. Z., Wang, P., Su, D. Q., Sun, M. M., et al. (2019). A College Genetic Analysis and QTL Mapping on Vertical Distribution of Pod Number in Soybean. *Soybean Sci.* 38, 360–370.
- Egli, D. B. (2005). Flowering, Pod Set and Reproductive Success in Soybean. *J. Agron. Crop Sci.* 4, 283–291. doi: 10.1111/j.1439-037X.2005.00171.x
- Endo, Y., and Ohashi, H. (2009). Diversification of Seed Arrangement Induced by Ovule Rotation and Septum Formation in Leguminosae. *J. Plant Res.* 122, 541–550. doi: 10.1007/s10265-009-0242-8
- Bai, F. and Demason, D. A. (2008). Hormone Interactions and Regulation of Pspk2: gus Compared With Dr5:Gus and Pid:Gus in Arabidopsis Thaliana. *Am. J. Bot.* 95, 133–145. doi: 10.3732/ajb.95.2.133
- Fang, D. M. (2008). Hormone interactions and regulation of Pspk2::GUS compared with Dr5:GUS and Pid:GUS in Arabidopsis thaliana. *Am. J. Bot.* 95, 133–145. doi: 10.3732/ajb.95.2.133
- Gao, J. Y., Liu, C. Y., Jiang, H. W., Hu, G. H., and Chen, Q. S. (2012). QTL Analysis of Pod Number per Plant in Soybean Under Multiple Locations. *China J. Oil Crop Sci.* 34, 1–7.
- Guo, M. L., Guo, T., Wang, Z. X., Zheng, W., Li, C. D., Zhao, H. H., et al. (2020). A Soybean Variety Henong 71 with High Yield Record in China and Its Super High Yield Cultivation Techniques. *Heilongjiang Agri. Sci.* 6, 139–141.
- Hu, G., Fan, J., Xian, Z., Huang, W., Lin, D., and Li, Z. (2014). Overexpression of Slrev Alters the Development of the Flower Pedicel Abscission Zone and Fruit Formation in Tomato. *Plant Sci.* 229, 86–95. doi: 10.1016/j.plantsci.2014.08.010
- Hyten, D. L., Choi, I. Y., Song, Q., Specht, J. E., Carter, T. E., Shoemaker, R. C., et al. (2010). A High Density Integrated Genetic Linkage Map of Soybean and the Development of a 1536 Universal Soy Linkage Panel for Quantitative Trait Locus Mapping. *Crop Sci.* 3, 960–968. doi: 10.2135/cropsci2009.06.0360
- Jiang, H., Li, Y., Qin, H., Li, Y., Qi, H., Li, C., et al. (2018). Identification of Major QTL Associated with First Pod Height and Candidate Gene Mining in Soybean. *Front. Plant Sci.* 9:1280. doi: 10.3389/fpls.2018.01280
- Jiang, Y. F., Lahlali, R., Karunakaran, C., Kumar, S., and Davis, A. R. (2015). Seed Set, Pollen Morphology and Pollen Surface Composition Response to Heat Stress in Field Pea. *Plant Cell Environ.* 11, 2387–2397.
- Kim, E., Kurdyukov, S., and Rose, R. J. (2009). Expression of the SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 (SERK1) gene is associated with developmental change in the life cycle of the model legume Medicago truncatula. *J. Exp. Bot.* 6, 1759–1771.
- Kurdyukov, S., Song, Y., Sheahan, M. B., and Rose, R. J. (2014). Transcriptional Regulation of Early Embryo Development in the Model Legume Medicago Truncatula. *Plant Cell Rep.* 33, 349–362. doi: 10.1007/s00299-013-1535-x
- Kwak, M., Velasco, D., and Gepts, P. (2008). Mapping Homologous Sequences for Determinacy and Photoperiod Sensitivity in Common Bean (Phaseolus Vulgaris). *J. Hered.* 99, 283–291.
- Lenser, T., Graeber, K., Cevik, ÖS., Adigüzel, N., Dönmez, A. A., Grosche, C., et al. (2016). Developmental Control and Plasticity of Fruit and Seed Dimorphism in Aethionema Arabicum. *Plant Physiol.* 172, 1691–1707. doi: 10.1104/pp.16.00838
- Li, C. D., Zou, J. N., Jiang, H. W., Yu, J. Y., Huang, S. Y., Wang, X. Y., et al. (2018). Identification and Validation of Number of Pod and Seed Related Traits QTL in Soybean. *Plant Breeding* 137, 730–745. doi: 10.1111/pbr.12635
- Li, D. M., Sun, M. M., Han, Y. P., Teng, W. L., and Li, W. B. (2010). Identification of QTL Underlying Soluble Pigment Content in Soybean Stems Related to Resistance to Soybean White Mold (Sclerotinia Sclerotiorum). *Euphytica* 172, 49–57. doi: 10.1007/s10681-009-0036-z
- Li, Y. Y., Li, R. C., Cheng, C. G., Zhao, Y. Y., Liu, C. Y., Qi, Z. M., et al. (2018). Meta and Overview Analysis of QTL Associated with Pod and Seed Traits and Candidate Gene Mining in Soybean (Glycine max L. Merrill). *J. Agri. Biotech.* 26, 1821–1833.
- Liang, H. Z., Yu, Y. L., Yang, H. Q., Zhang, H. Y., Dong, W., Li, C. Y., et al. (2012). Genetic Analysis and QTL Mapping of Pod-Seed Traits in Soybean Under Different Environments. *Sci. Agri. Sin.* 45, 2568–2579.
- Liew, L. C., Singh, M. B., and Bhalla, P. L. (2014). Unique and Conserved Features of Floral Evocation in Legumes. *J. Integr. Plant Biol.* 56, 714–728. doi: 10.1111/jipb.12187
- Liu, C. Y., Qi, Z. M., Han, D. W., Han, D., Shan, D., Jiang, H., et al. (2010). QTL Analysis of Yield Components on Soybean Under Different Environments. *J. Northeast Agr. Univ.* 41, 1–9.
- Liu, S. P., Xue, H., Zhang, K. X., Wang, P., Su, D., Li, W., et al. (2019). Mapping QTL Affecting the Vertical Distribution and Seed Set of Soybean Pods. *Crop J.* 7, 694–706.
- Lora, J., Hormaza, J. I., Herrero, M., and Gasser, C. S. (2011). Seedless Fruits and The Disruption of a Conserved Genetic Pathway in Angiosperm Ovule Development. *PNAS* 108, 5461–5465. doi: 10.1073/pnas.1014514108
- McCouch, S. R., Cho, Y. G., Yano, M., Paul, E., Blinstrub, M., Morishima, H., et al. (1997). Report on QTL Nomenclature. *Rice Genet. Newsl.* 14, 11–13.
- Mena-Alí, J. I., and Rocha, O. J. (2005). Effect of Ovule Position within the Pod on the Probability of Seed Production in Bauhinia unguolata (Fabaceae). *Ann. Bot.* 3, 449–455. doi: 10.1016/j.aquaculture.2007.12.029
- Mukhtar, F. A., and Coyne, D. P. (1981). Inheritance and Association of Flower, Ovule, Seed, Pod, and Maturity Characters in Dry Edible Beans (Phaseolus Vulgaris L.). *J. Am. Soc. Hortic. Sci.* 106, 713–719.
- Ning, H., Yuan, J., Dong, Q., Li, W., Xue, H., Wang, Y., et al. (2018). Identification of QTL Related to the Vertical Distribution and Seed-Set of Pod Number in Soybean (Glycine max L. Merrill). *PLoS One* 13:e0195830. doi: 10.1371/journal.pone.0195830
- Nolan, K. E., Sergey, K., and Rose, R. J. (2009). Expression of the Somatic Embryogenesis Receptor-Like Kinase1 (Serkl) Gene is Associated with Developmental Change in the Life Cycle of The Model Legume Medicago Truncatula. *J. Exp. Bot.* 60, 1759–1771. doi: 10.1093/jxb/erp046
- Pathak, A. K., Singh, S. P., Gupta, Y., Gurjar, A. K., Mantri, S. S., and Tuli, R. (2016). Transcriptional Changes During Ovule Development in Two Genotypes of Litchi (Litchi Chinensis Sonn.) with Contrast in Seed Size. *Sci. Rep.* 6:36304. doi: 10.1038/srep36304
- Pushpavalli, R., Zaman-Allah, M., Turner, N. C., Baddam, R., Rao, M. V., and Vadez, V. (2015). Higher Flower and Seed Number Leads to Higher Yield Under Water Stress Conditions Imposed During Reproduction in Chickpea. *Funct. Plant Biol.* 42, 162–174. doi: 10.1071/FP14135
- Qi, T., Huang, H., Song, S., and Xie, D. (2015). Regulation of Jasmonate Mediated Stamen Development and Seed Production by a bHLH-MYB Complex in Arabidopsis. *Plant Cell* 27, 1620–1633. doi: 10.1105/tpc.15.00116
- Qi, Z., Huang, L., Zhu, R., Xin, D., Liu, C., Han, X., et al. (2014). A High-Density Genetic Map for Soybean Based on Specific Length Amplified Fragment Sequencing. *PLoS One* 9:e104871. doi: 10.1371/journal.pone.0104871
- Radoeva, T., and Weijers, D. (2014). A Roadmap to Embryo Identity in Plants. *Trends Plant Sci.* 19, 709–716. doi: 10.1016/j.tplants.2014.06.009
- Safavian, D., Zayed, Y., Indriolo, E., Chapman, L., Ahmed, A., and Goring, D. (2015). Rnasilencing of exocyst genes in the stigma impairs the acceptance of compatible pollen in Arabidopsis. *Plant Physiol.* 169, 2526–2538.
- 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
- Seefried, W. F., Willmann, M. R., Clausen, R. L., and Jenik, P. D. (2014). Global Regulation of Embryonic Patterning in Arabidopsis by MicroRNAs. *Plant Physiol.* 165, 670–687. doi: 10.1104/pp.114.240846
- Shibles, R., Anderson, I. C., and Gibson, A. H. (1975). “Soybean” in *Crop Physiology: some Case Histories*. Ed L.T. Evans. (Cambridge: Cambridge University Press). 151–189. doi: 10.2136/sssaj1975.03615995003900050004x
- Smitha Ninan, A., Shah, A., Song, J., and Jameson, P. E. (2017). Differential Gene Expression in The Meristem and During Early Fruit Growth of Pisum Sativum L. Identifies Potential Targets for Breeding. *Int. J. Mol. Sci.* 18:428. doi: 10.3390/ijms18020428

- Song, J., Liu, Z., Hong, H., Ma, Y., Tian, L., Li, X., et al. (2016). Identification and Validation of Loci Governing Seed Coat Color by Combining Association Mapping and Bulk Segregation Analysis in Soybean. *PLoS One* 11:e0159064. doi: 10.1371/journal.pone.0159064
- Song, Q. J., Marek, L. F., Shoemaker, R. C., Lark, K. G., Concibido, V. C., Delannay, X., et al. (2004). A New Integrated Genetic Linkage Map of the Soybean. *Theor. Appl. Genet.* 109, 122–128. doi: 10.1007/s00122-004-1602-3
- Stasko, A. (2018). *Functional Gene Analysis of Resistance QTL Towards Phytophthora sojae on Soybean Chromosome 19*. Doctoral dissertation. Columbus, OH: Ohio State University.
- Su, D. Q., Jiang, S. T., Wang, J. J., Yang, C., Li, W. X., and Ning, H. L. (2019). Identification of Major QTL Associated with Agronomical Traits and Candidate Gene Mining in Soybean. *Biotechnol. Biotech. EQ.* 33, 1481–1493. doi: 10.1080/13102818.2019.1674691
- Tavares, L. C., Rufino, C., Brunes, A. P., Tunes, L. M., and Peske, S. T. (2013). Performance of Soybean Seeds Under Water Stress: yield and Physiological Quality of F1 Generation. *Ciência Rural*. 43, 1357–1363. doi: 10.1590/S0103-84782013000800003
- Tischner, T., Allphin, L., Chase, K., Orf, J. H., and Lark, K. G. (2003). Genetic of Seed Abortion and Reproductive Traits in Soybean. *Crop Sci.* 43, 464–473. doi: 10.2135/cropsci2003.0464
- Tsuwamoto, R., Yoko, S., and Takahata, Y. (2010). Arabidopsis EMBRYOMAKER Encoding an AP2 Domain Transcription Factor Plays a Key Role in Developmental Change From Vegetative to Embryonic Phase. *Plant Mol. Biol.* 73, 481–492. doi: 10.1007/s11103-010-9634-3
- Wang, S., Basten, C., and Zeng, Z. (2006). *Windows QTL Cartographer 2.5*. Raleigh: North Carolina State University.
- Wang, X. Z. (2008). *Inheritance and Stability Analysis and QTL Mapping of Yield Related Traits in Soybean*. Wuhan: Chinese Academy of Agricultural and Science. doi: 10.7666/d.Y1422503
- Wang, Y., Chai, C., Valliyodan, B., Maupin, C., Annen, B., and Nguyen, H. T. (2015). Genome-Wide Analysis and Expression Profiling of the Pin Auxin Transporter Gene Family in Soybean (*Glycine max* L. Merrill). *BMC Genomics* 16:951. doi: 10.1186/s12864-015-2149-1
- Wang, Z. (2004). *Construction of Soybean SSR Based Map and QTL Analysis of Important Agronomic Traits*. China: Guangxi University. doi: 10.7666/d.Y620581
- Xia, Y., Won, S., Du, X., Lin, P., Ross, C., La Vine, D., et al. (2010). Bulk Segregation Mapping of Mutations in Closely Related Strains of Mice. *Genetics*. 186, 1139–1146. doi: 10.1534/genetics.110.121160
- Xin, D. W., Qi, Z. M., Jiang, H. W., Zhang, Z. B., Zhu, R. S., Hu, J. H., et al. (2016). QTL Location and Epistatic Effect Analysis of 100-Seed Weight Using Wild Soybean (*Glycine soja* Sieb & Zucc) Chromosome Segment Substitution Lines. *PLoS One* 3:e0149380. doi: 10.1371/journal.pone.0149380
- Xue, Z. G., Zhang, X. M., Lei, C. F., Chen, X. J., and Fu, Y. F. (2012). Molecular Cloning and Functional Analysis of One Zeitrup Homolog Gmzt13 in Soybean. *Mol. Biol. Rep.* 39, 1411–1418. doi: 10.1007/s11033-011-0875-2
- Yang, X. C., and Hwa, C. M. (2008). Genetic Modification of Plant Architecture and Variety Improvement in Rice. *Heredity* 101, 396–404. doi: 10.1038/hdy.2008.90
- Yang, Z., Xin, D. W., Liu, C. Y., Jiang, H. W., Han, X., Sun, Y. N., et al. (2013b). Identification of QTL for Seed and Pod Traits in Soy_x0002_Bean and Analysis for Additive Effects and Epistatic Effects of QTL Among Multiple Environments. *Mol. Genet. Genomics* 288, 651–667. doi: 10.1007/s00438-013-0779-z
- Zhang, D., Cheng, H., Wang, H., Zhang, H., Liu, C., and Yu, D. (2010). Identification of Genomic Regions Determining Flower and Pod Numbers Development in Soybean (*Glycine max* L. Merrill). *J. Genet. Genomics*. 37, 545–556. doi: 10.1016/S1673-8527(09)60074-6
- Zhang, W. Y., Wang, J. X., Zong, C. M., Fu, C. X., Wang, M. Q., Sun, Y. J., et al. (2012). QTL Analysis of Seed and Pod Traits in Soybean. *Soybean Sci.* 31, 193–197.
- Zhao, S. P., Lu, D., Yu, T. F., Zheng, W. J., Zhang, S. X., Chai, S. C., et al. (2017). Genome-wide Analysis of the YABBY Family in Soybean and Functional Identification of GmYABBY10 Involvement in High Salt and Drought Stresses. *Plant Physiol. Biochem.* 119, 132–146. doi: 10.1016/j.plaphy.2017.08.026
- Zhou, R., Chen, H. F., Wang, X. Z., Zhang, X. J., and Wu, J. S. (2009). QTL Analysis of Yield, Yield Components and Lodging in Soybean. *Acta Agronom. Sin.* 35, 821–830. doi: 10.3724/SP.J.1006.2009.00821
- Zhou, X. A., Wang, X. Z., Cai, S. P., Jun, W. X., Sha, A. H., Qiu, D. Z., et al. (2005). Relation of Three-seed and Four-seed Pods with Yield of RIL in Soybeans. *China. J. Oil Crop Sci.* 27, 22–25.

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 © 2021 Li, Jiang, Li, Liu, Qi, Wu, Zhang, Hu, Zhu, Guo, Wang, Zheng, Zhang, Zhao, Wang, Shan, Xin, Luan and Chen. 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.



Establishment and Application of Ligation Reaction-Based Method for Quantifying MicroR-156b

Yuxuan He, Likun Long, Wei Yan, Liming Dong, Wei Xia, Congcong Li and Feiwu Li*

Institute of Agricultural Quality Standard and Testing Technology, Jilin Academy of Agricultural Sciences, Changchun, China

OPEN ACCESS

Edited by:

Xianzhong Feng,
Northeast Institute of Geography
and Agroecology, Chinese Academy
of Sciences (CAS), China

Reviewed by:

Fabrizio Cillo,
Institute for Sustainable Plant
Protection, National Research Council
(CNR), Italy
Yinan Yuan,
Michigan Technological University,
United States

*Correspondence:

Feiwu Li
lifeiwu3394@sina.com

Specialty section:

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

Received: 14 October 2021

Accepted: 25 November 2021

Published: 14 December 2021

Citation:

He Y, Long L, Yan W, Dong L,
Xia W, Li C and Li F (2021)
Establishment and Application of
Ligation Reaction-Based Method for
Quantifying MicroR-156b.
Front. Plant Sci. 12:794752.
doi: 10.3389/fpls.2021.794752

Microribonucleic acids (miRNAs) play significant roles in the regulation of biological processes and in responses to biotic or abiotic environmental stresses. Therefore, it is necessary to quantitatively detect miRNAs to understand these complicated biological regulation mechanisms. This study established an ultrasensitive and highly specific method for the quantitative detection of miRNAs using simple operations on the ground of the ligation reaction of ribonucleotide-modified deoxyribonucleic acid (DNA) probes. This method avoids the complex design of conventional reverse transcription. In the developed assay, the target miRNA miR156b was able to directly hybridize the two ribonucleotide-modified DNA probes, and amplification with universal primers was achieved following the ligation reaction. As a result, the target miRNA could be sensitively measured even at a detection limit as low as 0.0001 amol, and differences of only a single base could be detected between miR156 family members. Moreover, the proposed quantitative method demonstrated satisfactory results for overexpression-based genetically modified (GM) soybean. Ligation-based quantitative polymerase chain reaction (PCR) therefore has potential in investigating the biological functions of miRNAs, as well as in supervising activities regarding GM products or organisms.

Keywords: MiRNAs, ribonucleotide-modified DNA probe, ligation reaction, quantitative method, MiR156b

INTRODUCTION

Micro ribonucleic acids (miRNAs) are a group of endogenous, non-coding, single-stranded small molecular RNA (18–25 nt in length) that participate in the post-transcriptional regulation of gene expression. miRNAs are emerging as novel biomarkers, with profound implications for research into the responses to abiotic (Aslam et al., 2020; Singroha et al., 2021) or biotic stress (Chopperla et al., 2020; Šečić et al., 2021), development and differentiation in cellular behavior (Krol et al., 2010; Miao et al., 2020), and multiple plant signaling pathways (Curaba et al., 2014). Since their first discovery, several miRNAs have been identified in a wide range of species. Following the identification of the functions of miRNAs in crops, a large number of transgenic crops related with miRNA have been developed and commercialized worldwide (Agrawal et al., 2015; Bally et al., 2020). Among the miRNAs in plant, microRNA156 (miR156) and its target mRNAs has been widely observed as a major regulators for crop development (Cao et al., 2015) and abiotic stresses (Visentin et al., 2020). The miR156 expresses at the highest accumulation at seedling stage (Miao et al., 2019). Moreover, the expression of the other miRNAs, such as miR159, miR33 might have a effect on miR156 level (Guo et al., 2017). In the whole growth stage of grapevine,

miR156b/c/d exhibited typical temporal spatial-specific expression levels (Wang et al., 2016). The Arabidopsis AtmiR156b gene overexpressed in GM Brassica napus resulted in seed lutein and beta-carotene accumulation, these achievement suggest that miR156b have a potential application in plant breeding for enhancing carotenoid production (Wei et al., 2010). To further explore the physiological functions of miRNA or detect the content in crop genomes, specific quantitative detection methods for miRNAs that have high sensitivity, efficiency, and reliability are urgently needed for miRNA mechanism applications (Papadopoulou et al., 2020). In early miRNA studies, northern blotting (De la Rosa and Reyes, 2019) and microarray (Li and Ruan, 2009) assays were conventionally used as “gold standard” methods to identify miRNA expression. However, these complex and relatively insensitive methods cannot overcome the complications associated with the unique characteristics of miRNAs such as their low abundances in tissue, short lengths, and the high sequence similarities among miRNA families. To resolve these problems, a series of cooperative amplification programs have been reported for miRNA detection. Locked nucleic acid (LNA)-modified probes with bicyclic high-affinity RNA analogs, which are rigid structures formed by linking the 2' and 4' carbons on the ribose, greatly improve the hybridization efficiency of northern blotting and are sensitive to miRNA determination (Shu et al., 2019). The Agilent Microarray SurePrint technology is advantageous owing to its low RNA input requirements (~100 ng) and eliminates bias by omitting the fractionation or amplification steps (Dell'Aversana et al., 2017). However, many solid phase hybridization assays require sophisticated fabrication or expensive capture probe modifications (Zou et al., 2017). These issues increase their costs and restrict their extensive application. Modified miRNA detection methods based on nucleic acid amplification are undoubtedly the most attractive and prominent tools for miRNA analysis and include polymerase chain reaction (PCR), ligase chain reaction (LCR) (Zhang et al., 2013), rolling circle amplification (RCA) (Cheng et al., 2009; Liu et al., 2020), and isothermal exponential amplification reaction (IEXPAR) (Jia et al., 2010). For instance, with the design of a new hairpin/deoxyribonucleic acid (DNA) ring ternary probe in RCA, the target miRNA can bind the probes through toehold-mediated strand displacement to form ternary structures and generate many repeated metal ion-dependent deoxyribozyme (DNAzyme) sequences. Finally, fluorescently quenched hairpin signal probes can be cyclically cleaved by the DNAzyme sequences, which drastically enhances the fluorescence recovery of the target miRNA (Liu et al., 2019). Tian et al. (2019) rationally combined RCA with efficient loop-mediated isothermal amplification (LAMP), which directly templates the ligation of a padlock probe to trigger the RCA reaction. Thus, they significantly improved the amplification efficiency and sensitivity owing to the double stem-loop DNAs with functional sequences that were generated by RCA-produced DNA templates.

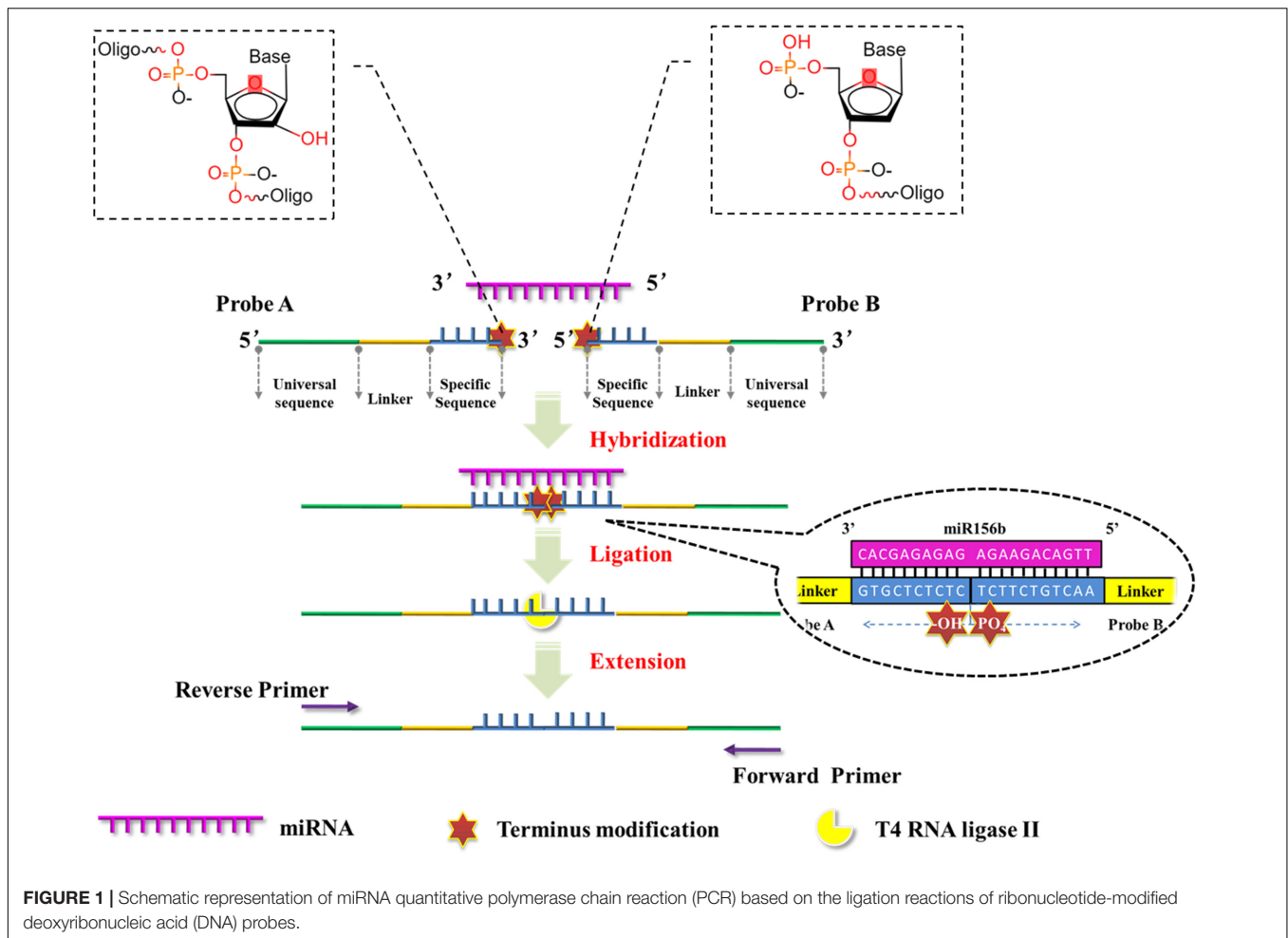
This study proposes the development of a facile and ultrasensitive method for miRNA detection by elegantly integrating sensitivity and specificity. Here, a miRNA quantification assay is achieved through real-time PCR

integrated ligation reactions. As outlined in the schematic representation in **Figure 1**. After determining the target miRNA to be detected, the two DNA probes in this assay contain a specific sequence (blue), a stuffer sequence (yellow) in between and a universal sequence designed as amplification primer (green). Both of the specific sequences are complementary to the half-sequence of target miRNA at the 3'- and 5'-termini, respectively. Then, the 3'-termini of probe A is modified by two nucleotides and the 5'-termini of probe B is modified by a phosphate group. Therefore, the modified groups are immediately adjacent to each other following the hybridization of the probes with the half-sequence of target miRNA. As T4 RNA ligases specifically catalyze the formation of a 3'-5'-phosphodiester bond between a 3'-hydroxyl group and a 5'-phosphoryl group in three nucleotidyl transfer steps (Zhuang et al., 2012), and T4 RNA ligase II exhibits significantly higher ligating activity and specificity for double stranded RNA than T4 RNA ligase I (Yin et al., 2003; Cheng et al., 2009). T4 RNA ligase II is employed to link the 3'-hydroxyl and 5'-phosphoryl groups in adjacent probes and thus form ligation production, which is used as an initial template for the next real-time PCR amplification with the universal sequences. The PCR products with fluorescence or SYBR Green I signal can then be detected directly. The detection limit of the proposed method was determined and its specificity was investigated with respect to its ability to distinguish miRNAs from their corresponding pre-miRNAs, and from homogeneous miRNAs in the same family. The robustness and feasibility of the method were investigated as a tool for genetically modified (GM) content analysis.

MATERIALS AND METHODS

Plant Materials and Extract Preparation

All plant materials were kindly provided by Jilin Academy of Agricultural Sciences, Changchun, China, and germinated in soil (pH 7.19; soil: water 1: 4) under a photoperiod of 16 h of daylight at 23°C. A 181-bp stem-loop fragment of the *GmmiR156b* (GenBank No. NR048600.1) precursor was amplified using DNA samples obtained from the soybean cultivar Williams 82 and subcloned into the *pTF101.1-35S* vector, which carried the selection marker *Bar* gene and the cauliflower mosaic virus 35S promoter that could enhance the expression of miR156b. The recombinant construct *pTF101.1-miR156b* was then integrated into *Agrobacterium tumefaciens* strain EHA101 for cotyledon-node transformation (Flores et al., 2008). After the regenerated plants were transplanted into a greenhouse, glufosinate (150 mg/L) screening and LibertyLink strip (EnviroLogix Inc., Portland, ME, United States) detection were used to identify transgenic plants. Herbicide-resistant T4 generations containing a single copy number of the target sequence were employed as miR156b GM soybean event for further experimental analyses. In this GM soybean, the expression level of miR156b has been considered by the conventional stem-loop qPCR, which further confirmed the reliability



of the plant materials. Total RNA and DNA samples were extracted from eight-week-old transgenic and wild-type (WT) soybean seedlings. All procedures were performed in accordance with the EasyPure Plant RNA kit (Transgen Biotech, Beijing, China) and Dneasy Plant Mini Kit (Qiagen, Hilden, Germany). The extracted concentration and quality were calculated using absorbance measurements with an ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, United States) at wavelengths of 260 and 280 nm. All the extracted samples were then stored at -80°C for further research.

Design of Probes and Primers for miR156 RNA Detection

All the miRNA sequences in the plant can be obtained from the online database PmiREN.¹ The miR156 family comprises of typical miRNAs that are considered to have an explicit function applied in some transgenic soybeans; in this assay, the miR156b gene from the miR156 family was used as the model target. The structures of probes A and B are elucidated schematically in **Figure 1**. The specific sequences

(blue) of probes A and B were designed to be complementary to half of the target miRNA sequences at the 3'- and 5'-terminals, respectively. A stuffer sequence (yellow) in between links up the specific sequence with the universal sequence (green). Probe A was modified with two nucleotides at its 3'-terminus, while probe B was modified with a phosphate group at its 5'-terminus. As the chemical features of the modified nucleotides, 3'-hydroxyl in probe A and 5'-phosphoryl groups in probe B (**Figure 1**) are immediately adjacent to each other following the DNA probes hybridize with the target miRNA, and the modified structure facilitates the formation of a 3'-to-5'-phosphodiester bond catalyzed by T4 RNA ligase II. Moreover, probe A-N has the same sequences as probe A except for the presence of 3'-hydroxyl groups in the two 3'-terminal ribonucleotides. Probe A-N was employed as an experimental control to verify the role of terminus modifications in the ligation reaction. In order to test the specificity of this proposed method in miRNA discrimination, for another miR156 member (miR156g), probes A-g and B-g were also designed in the same way following the replacement of specific sequences (**Table 1**).

The forward and reverse primers for PCR amplification were designed according to the universal sequences in probes

¹<https://www.pmiREN.com/search>

TABLE 1 | The sequences of the primers and probes in the ligation based quantitative polymerase chain reaction (PCR) amplification.

ID	Sequences (5'-3')
Probe A	GGGATACTGGAACCTGATGATGACTAACTAACTACCTG GGCTGAT GTGCTCTCrTrC
Probe B	PO ₄ - TCCTCTGTCAA CCTTTGCTTT ACTACTCTCACACACTCTATGCTTGCTACCGTCG
Probe A-N	GGGATACTGGAACCTGATGATGACTAACTAACTACCTG GGCTGAT GTGCTCTCTC
Probe A-g	GGGATACTGGAACCTGATGATGACTAACTAACTACCTG GGCTGAT GGAGCTCCrCrT
Probe B-g	PO ₄ - TCACTCCAAT CCTTTGCTTT ACTACTCTCACACACTCTATGCTTGCTACCGTCG
Forward primer-1	CGACGGTAGCAAGCATAGAGTGTG
Reverse primer-1	GGGATACTGGAACCTGATGATGAC
Forward primer-2	CGACGGTAGCAAGCATAGAGTGTGTGAGAG
Reverse primer-2	GGGATACTGGAACCTGATGATGACTAACTAACTACC
Forward primer-3	CGACGGTAGCAAGCATAGAGTGTGTGA
Reverse primer-3	GGGATACTGGAACCTGATGATGACTAAC
Pre-Gma-MIR156b	GTGATGTGAGATATCTCATGTTGACAGAAGAGAGAGAGCACAACCCGGGAATGGCTAA AGGAGTCTTTGCCTTTGTTGGGAGTGTGCCCTCTCTTCTCTGTCATCATCACATTCACA TGCTTTGC
Forward primer for pre-miR156	5'-TAATACTGACTCACTATAGGGAAGAGAGAGAGCACAACCcggaatggctaaggagtc-3'
Reverse primer for pre-miR156	5'-GACAGAGGAAGAGAGGGCACACTCCCAACAAAGGCAAAGactccttagcattcccg-3'

The sequences indicated with bold italics were the complementary sequence to the target miRNA.

The bold in forward primer for pre-miR156 was the specific sequence of pre-miRNA.

The box was the stuffer sequence.

B and A, respectively. For the optimization of the ligation-based qPCR, primer pairs with different lengths were designed and orthogonally combined to screen the optimal primers for the amplification of the ligation product. The sequences of the primers and probes are shown in Table 1. All high-performance liquid chromatography-purified miRNAs, polyacrylamide gel electrophoresis-purified ribonucleotide-modified DNA probes, and primers were synthesized by Shanghai Sangon Biological Engineering Technology (Shanghai, China).

Experimental Procedure of Ligation-Based Quantitative Polymerase Chain Reaction Assay

The mixture contained 2 nM each of template miRNA and probes A and B. The procedure was initially performed at 65°C for 8 min. In this step, both the DNA probes identify and hybridize with the target miRNA. Then, 2 U of T4 RNA ligase II and 10× ligation buffer (New England Biolabs, Shanghai, China) were added to the hybridization product, and a total final volume of 50 µl was incubated at 37°C for 1 h to perform the ligation reaction; the products were then immediately placed on ice until it cooled to room temperature (~15 min).

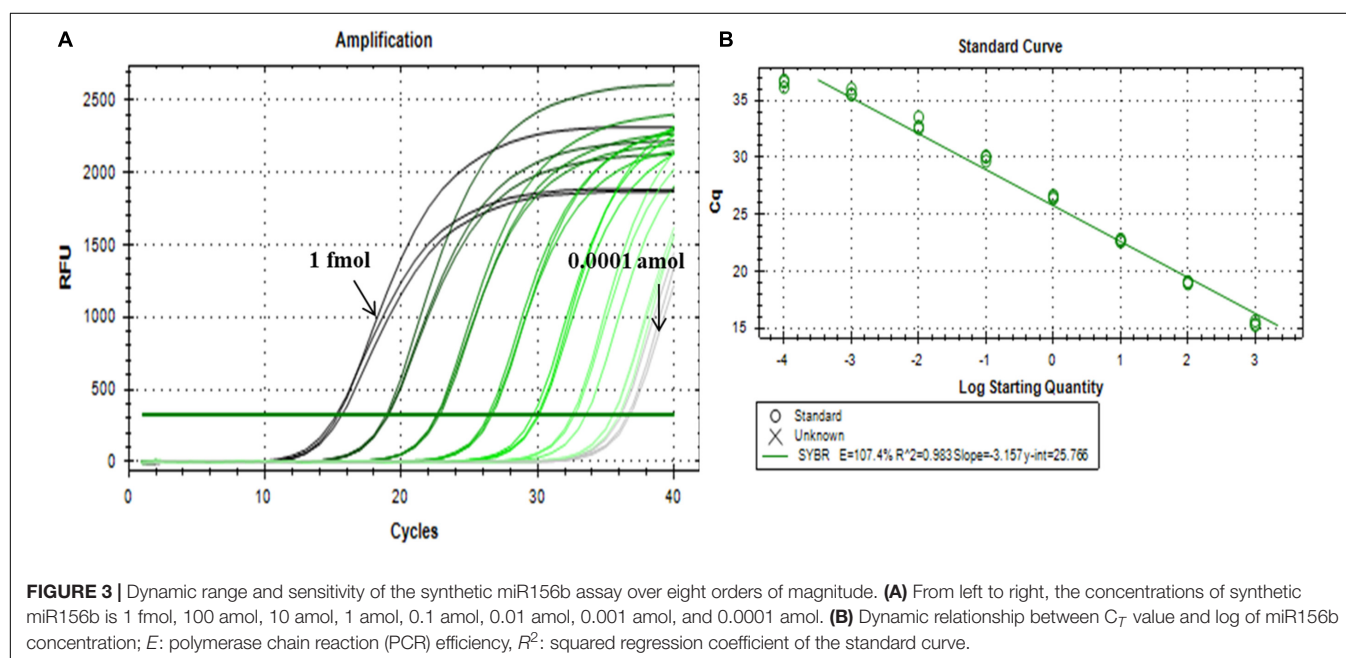
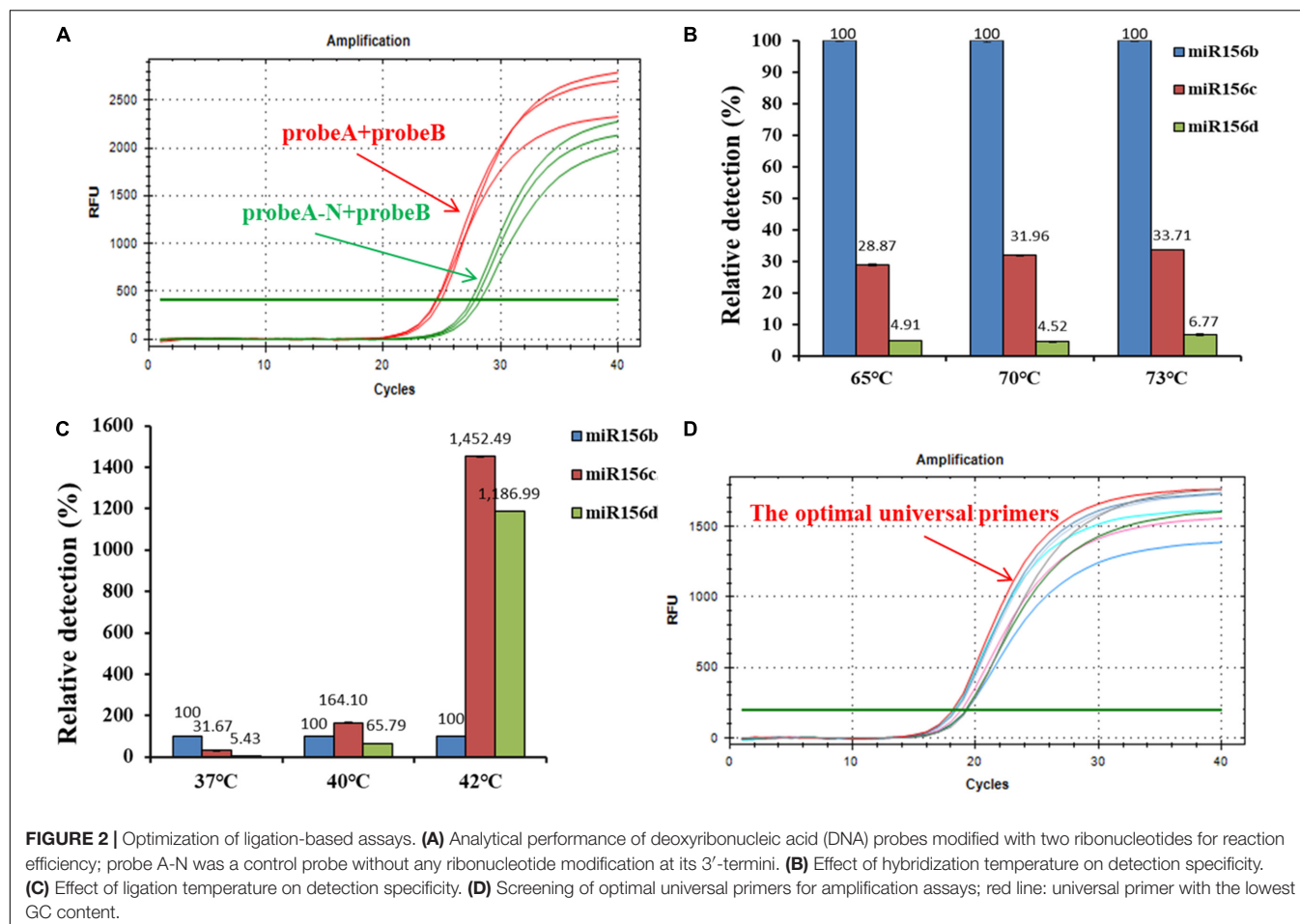
The FastStart Universal SYBR® Green Master (Roche, Mannheim, Germany) was used for real-time quantitative PCR assays. The quantitative PCR assays for amplifying the target miRNA were performed with a 20-µl final volume containing 2× SYBR Green Mastermix (ROX), 0.2-µM forward and reverse primers, 5 µl of ligation production, and 4.2 µl of diethylpyrocarbonate-treated water (DEPC-H₂O; Takara Biotechnology, Changchun, China). The reactions were incubated at 95°C for 5 min on a Bio-Rad CFX96 Real-time Thermal Cycler (Bio-Rad, Hercules, CA, United States), followed

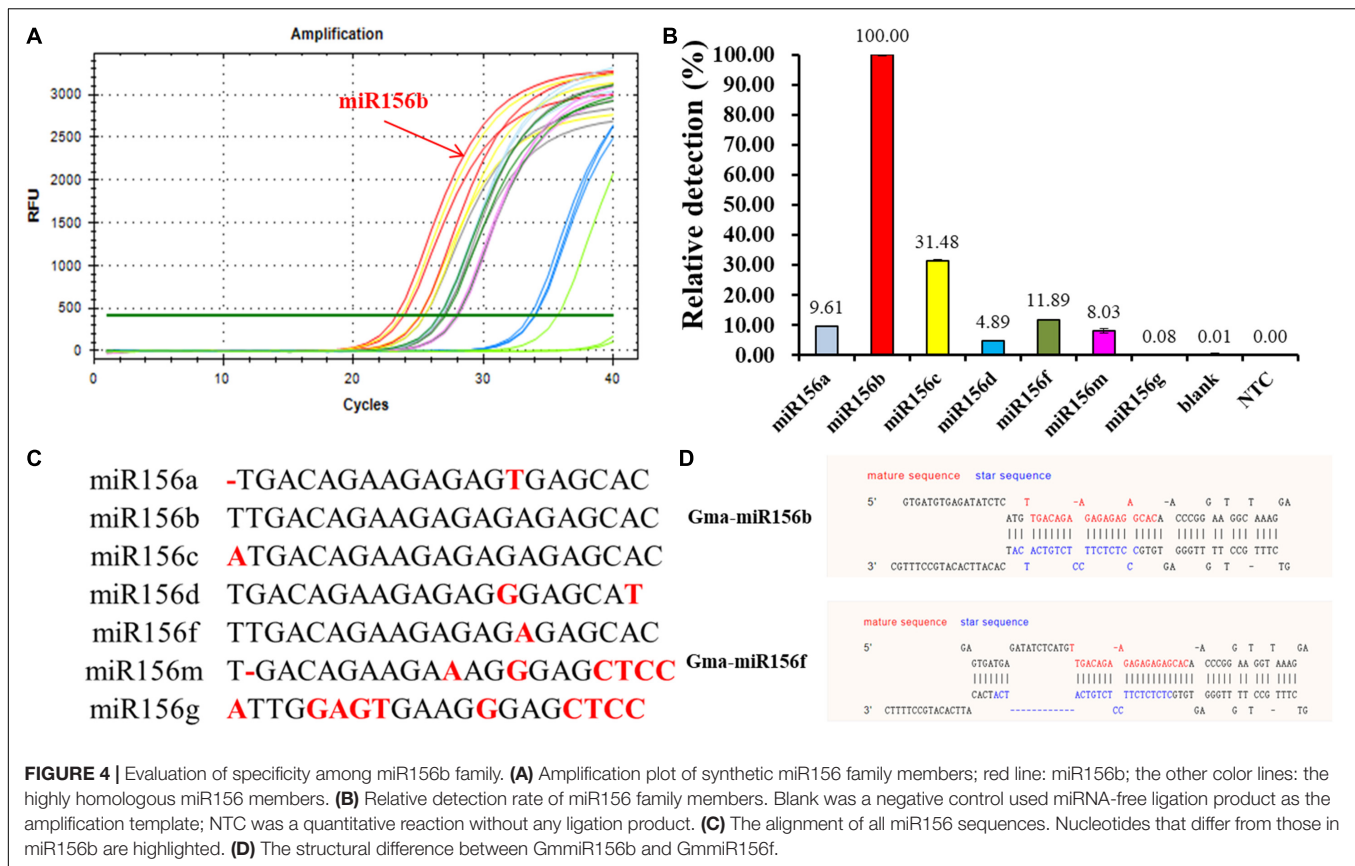
by 40 cycles of 95°C for 10 s, 60°C for 30 s, and then 72°C for 30 s. All reactions were performed in triplicate.

Preparation of Pre-miR156b *in vitro* Transcription Assays

The precursor of miR156b, pre-miR156b, was prepared using an *in vitro* transcription reaction. First, the forward and reverse primers for pre-miR156b (pre-FP and pre-RP) were designed based on the sequences of pre-GmmiR156b,² in which the 20 bases located at the 3'-terminus of pre-FP complemented each other with the 20 bases located at the 3'-terminus of pre-RP. A mixture containing 50 pmol of primer and 10 µl of Klenow buffer (Fermentas, Shanghai, China) was incubated at 75°C for 5 min and cooled on ice. After the 20 complementary bases hybridized completely with each other, deoxynucleotide triphosphate (dNTP; 250 µM final concentration; Takara Biotechnology, Changchun, China), Klenow buffer, and 5 U of Klenow DNA polymerase (exo-; Fermentas, Shanghai, China) were added and replenished with DEPC-H₂O (Takara Biotechnology, Changchun, China) to a final volume of 20 µl. Following incubation at 37°C for 1 h, the 3'-terminus sequences of both pre-FP and pre-RP generated double-stranded DNA (dsDNA) by extension reactions. Then, the reaction mixture was heated at 75°C for 20 min to inactivate the Klenow DNA polymerase. The complete sequence consists of the T7 promoter, GGGA (guanine-guanine-guanine-adenine) spacer, and a pre-miR156b specific sequence from the 5'- to 3'-terminus sequence of this dsDNA. The *in vitro* transcription reaction for pre-miR156b was performed in a final volume of 50 µl containing 20 µl of dsDNA, 30 µl of *in vitro* transcription buffer, 100 U of ribonuclease inhibitor

²<https://www.pmiren.com/singlemirna?Accession=PmiREN006720>





(Takara Biotechnology, Changchun, China), and 80 U of T7 RNA polymerase (Fermentas, Shanghai, China). The reaction mixture was incubated at 37°C for 4 h to amplify pre-miR156b, and then digested with 5 U of RNase-Free DNase I (Takara Biotechnology, Changchun, China). Finally, the above components were purified with an RNA cleanup and concentration kit (Shenggong, Shanghai, China) for 5 min. The concentration of the final pre-miR156b product was determined by analyzing its absorption at 260 nm with an ND-1000 spectrophotometer (Thermo Fisher Scientific Inc., Wilmington, DE, United States).

Data Processing Methodology

For data analysis, a 10-fold dilution series of ligation products was used as a template for quantitative PCR assays. This was done to generate a plot of log copy numbers of miRNAs vs. the corresponding threshold of cycle (C_T). The concentration of the target miRNA was detected through its linearity with C_T , which spanned at least seven orders of magnitude. Slope and PCR amplification efficiency (E) were calculated according to the following equation: $\text{slope} = -(1/\log E)$. The relative detection rate of other miR156 family members was calculated as $1/2^{\Delta C_T}$, in which $\Delta C_T = C_{T, \text{other miRNA}} - C_{T, \text{miR156b}}$. The relative detection rate of miR156b was defined as 100% (Zhang et al., 2011).

RESULTS AND DISCUSSION

Analytical Performance of Deoxyribonucleic Acid Probes Modified With Two Ribonucleotides for Reaction Efficiency

The high specificity of T4 RNA ligase II is dependent on the 3'-hydroxyl and 5'-phosphoryl groups that are adjacent to either a duplex RNA or an RNA-DNA hybrid. Thus, the effect of the modification of two nucleotides on reaction efficiency was examined in the miR156b detection assay. After either probe A/probe B or probe A-N/probe B was hybridized with the target miRNA and ligated by the T4 RNA ligase II, the ligation product was used as an initial template for PCR amplification. As seen in the real-time fluorescence signal in **Figure 2A**, templates with concentrations as low as 2 nM could be accurately detected by the probe A/probe B set in the ligation step, with C_T values ranging from 24.47 to 24.91 in triplicate. However, under the same experimental conditions, the ligation product using probe A-N/probe B had a C_T value ranging from 27.48 to 28.25 after PCR amplification was completed. These results demonstrate that although T4 RNA ligase II could directly catalyze DNA nick ligation in a way, ribonucleotide modification at the 3'-terminus greatly improved catalytic efficiency, and was more conducive to ligation-based PCR amplification.

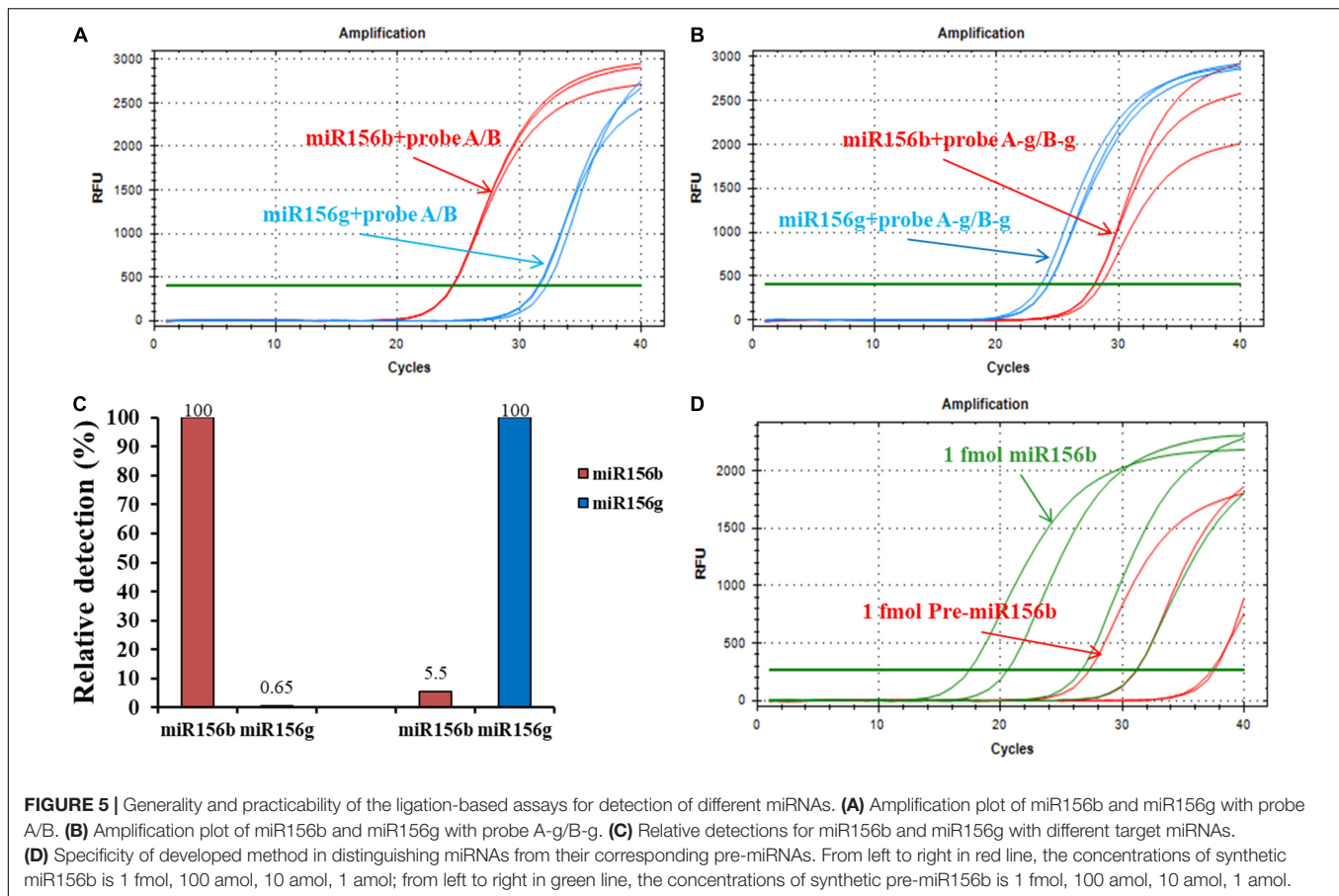


FIGURE 5 | Generality and practicability of the ligation-based assays for detection of different miRNAs. **(A)** Amplification plot of miR156b and miR156g with probe A/B. **(B)** Amplification plot of miR156b and miR156g with probe A-g/B-g. **(C)** Relative detections for miR156b and miR156g with different target miRNAs. **(D)** Specificity of developed method in distinguishing miRNAs from their corresponding pre-miRNAs. From left to right in red line, the concentrations of synthetic miR156b is 1 fmol, 100 amol, 10 amol, 1 amol; from left to right in green line, the concentrations of synthetic pre-miR156b is 1 fmol, 100 amol, 10 amol, 1 amol.

Optimization of Ligation Reaction of Ribonucleotide-Modified Deoxyribonucleic Acid Probes

The ability to distinguish similar miRNAs is essential for miRNA detection. The central goal of the developed ligation-based qPCR assay was to establish the appropriate reaction conditions to improve the specificity and reduce the unexpected ligation and hybridization with non-target miRNAs. To ensure high ligation reaction efficiency, the hybridization and ligation temperatures were optimized. In this ligation reaction, miR156b, miR156c, and miR156d were selected as experimental templates based on the similarity of the sequences. The results of the real-time fluorescence signal (Figures 2B,C) show that the hybridization temperature (from 65 to 73°C) had no consequence on the specificity to identify target miR156b, whereas the specificity of the ligation reaction decreased with increasing ligation temperature (from 37 to 42°C). When the ligation temperature was >40°C, the fluorescence signals of interference generated by miR156c were significantly enhanced compared to those of miR156d and miR156b. When the ligation temperature was set at 40 or 42°C, with equivalent molar amounts of the miRNA template, the cycle value of miR156b amplification was later than that at 37°C (Supplementary Figure 1). Data indicate that the initial amount of miR156 used for amplification was reduced in the real-time PCR assay. Based on the above mentioned

comprehensive experimental results, 37°C was considered as the optimal ligation temperature to ensure detection specificity.

The amplification efficiency of quantitative real-time PCR depends on the oligonucleotide primer sequences. In order to produce well-defined real-time fluorescence signals, several combinations of forward and reverse primers with different lengths were designed for PCR amplification. As depicted in Figure 2D and Supplementary Table 1, real-time fluorescence curves were generated by the three primer pair combinations. The C_T values of target miR156b ranged from 18.15 to 19.33. Although there was no significant difference between the experimental data of each group, the primer pair with the smallest C_T values was still employed as the optimal universal primer pair (forward primer: 5'-CGACGGTAGCAAGCATAGAGTGTGTGA-3'; reverse primer: 5'-GGGATACTGGAACCTGATGATGACTAACTAACTACC-3') for subsequent experiments. As the GC content was closely related to the annealing temperature, it is speculated that the lower GC content can promote the binding of primer and target template to a certain extent.

Dynamic Range and Sensitivity

Using the systematically optimized reaction parameters, a standard curve for evaluating the dynamic range and sensitivity of the ligation-based qPCR method was established *via* the

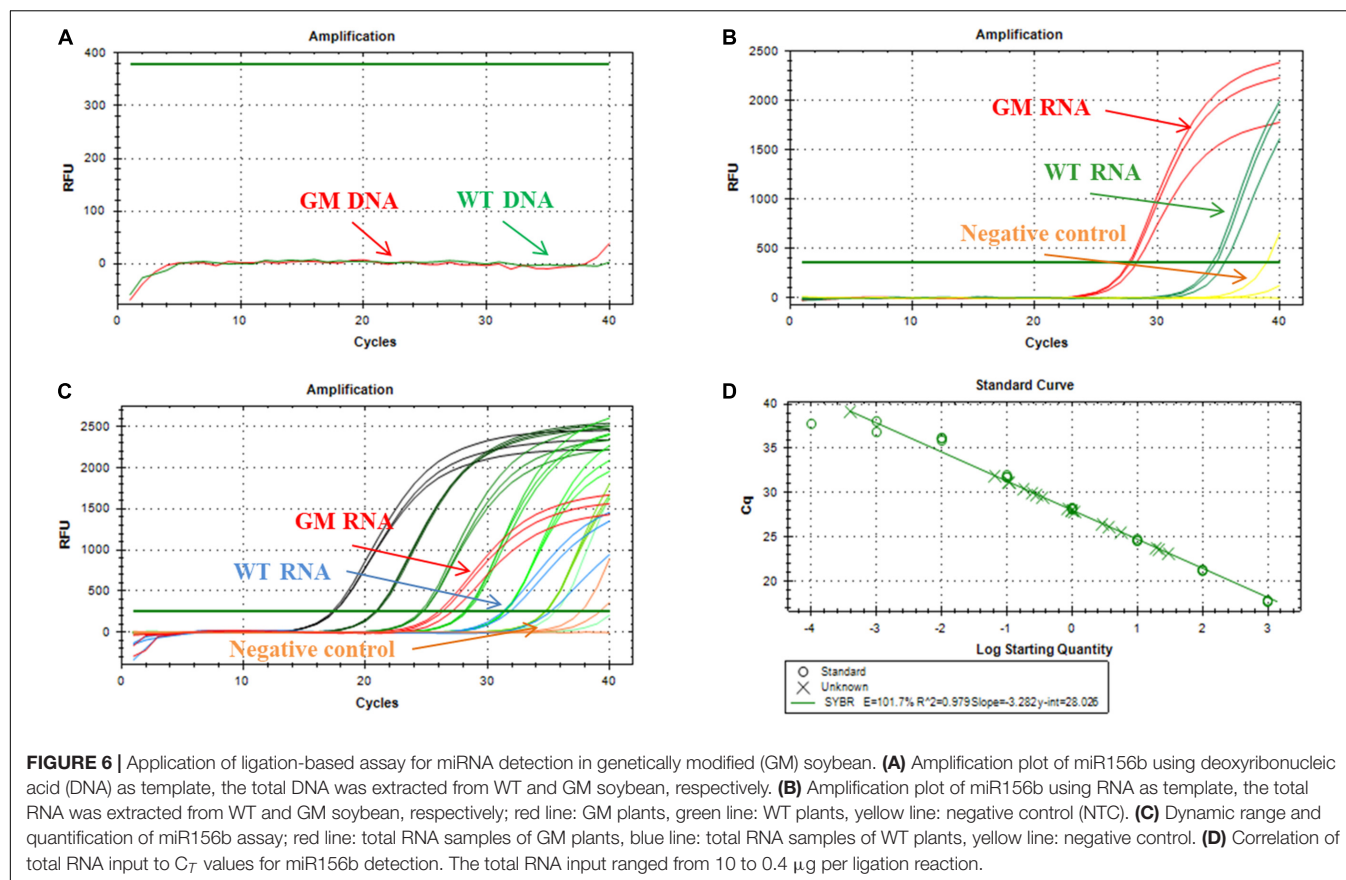


TABLE 2 | The comparison among ligation-based polymerase chain reaction (PCR) and the other similar methods for miRNA detection.

Methods	Detection limit	Analysis time	Advantages	Disadvantages	References
Ligation-based PCR	0.1 zmol	2.2 h	Simple probe design and easy to operate, good sensitivity and specificity	Special enzymes are needed, the result is affected by ligation efficiency	This work
Stem-loop RT-PCR	0.2 fmol	3 h	Good sensitivity and accuracy	Time consuming, complicated probe design	Mei et al., 2012
RCA assay	1.51 fmol	>10 h	Good efficiency of signal amplification, no PCR needed	Time consuming, complicated probe design, non-specific amplification of background	Liu et al., 2019
Hybridization chain reaction	10 pmol	>1 day	Better detectability, no enzyme is required	Cost consuming, specialized equipment needed	Zou et al., 2017
Isothermal exponential amplification reaction	0.1 zmol	<1 h	Good efficiency of signal amplification, no modified DNA probes is required, good sensitivity and specificity, low-cost	Multiple enzymes, specific nicking-enzyme recognition site and more condition optimization are required.	Jia et al., 2010

10-fold series dilution of ligation products as quantitative templates, spanning over at least seven orders of magnitude (miR156b inputs in the range of 1 fmol–0.0001 amol). This calibration method is superior to the delta threshold of cycle (C_T) method in terms of its detection limits (Taverniers et al., 2004). The quantitative PCR assay targeting miR156b further demonstrated an excellent linearity between the C_T value and the log of the target concentration in the range of 1 fmol–0.0001 amol. The correlation equation was $C_T = -3.157 \log$

$C_{miRNA} + 25.766$. The proposed ligation-based quantitative system showed a corresponding PCR efficiency of 107.4%, and the squared regression coefficient (R^2) of the standard curve was 0.983 for the three replicates (Figure 3). These results show that there was a good linear relationship between the target substance and PCR C_T value, over a wide range of concentrations. Notably, miR156b could be detected at concentrations as low as 0.0001 amol, demonstrating that the developed quantitative method has good sensitivity.

Specificity Study

The miR156 family contains seven highly homologous miRNAs with only one to a few mismatched nucleotides. To evaluate the specificity of the proposed qPCR method for miRNA detection, probes A and B were selected to verify whether the said method could identify other miR156 family members with different sequences. As shown in **Figure 4**, miR156b was identified, and each miR156 sequence obtained a different relative detection rate. Compared to miR156b, both miR156c and miR156f differed by only a single adenine nucleotide; the single-base nucleotide in miR156c was located at the 3'-terminus of probe B. Non-specific detection (31.48%) was higher than that of miR156f (11.89%), whose mature sequence was consistent with miR156b except for the star sequence (**Figures 4C,D**). This difference may further affect the secondary structure. The analytical data stored in the online database PmiREN (see text footnote 1) also confirms this aspect of miRNA. These results demonstrate that the secondary structure can affect specificity by influencing the binding between the probe and target miRNA to some extent. Both miR156a and miR156d had a mismatched nucleotide at the same position on the 3'-terminus of probe A, but another differential nucleotide in miR156a was located at the 5'-terminus of the mature miRNA sequence, while that of miR156d was at the 3'-terminus of the mature miRNA sequence. The higher non-specific detection determined for miR156a (9.61%) indicates that the method's specificity was affected not only by the number of differential nucleotides, but also by the location. In addition, two negative controls were established to verify the effect of the ligation between DNA probes on the fluorescence signals, one was a quantitative reaction without any ligation product (NTC), the other was a negative control used for the miRNA-free ligation product as the amplification template (blank). Neither of the negative controls generated notable fluorescence signals. Therefore, it can be inferred that the two DNA probes ligate each other in the absence of the target miRNA to produce a minimal effect on the quantitative results. Collectively, the proposed ligation-based qPCR method can characterize with a high specificity, discriminate differences as small as a single nucleotide, and detect secondary structural differences among miRNA targets.

Practicability Validation for miRNA Detection

A different miRNA (miR156g) was randomly selected to verify the universal practicability of the developed ligation-based quantitative PCR methods. To this end, specific probes A-g and B-g were designed with modified ribonucleotides containing the target-specific sequence that is complementary to the half sequence of miR156g. The stuffer sequence and universal sequences for PCR amplification were the same as those for probes A and B. Distinct results can be observed in **Figures 5A–C**, where equivalent amounts of miR156b and miR156g with different probes hybridized with respective or adverse probes produced well-defined signals. When using miR156b-specific probes (probe A and probe B) to perform the

ligation reaction, and calculate the relative detection level with the formula:

$$\Delta C_T = C_{T, \text{miR156g}} - C_{T, \text{miR156b}}$$

The non-specific detection rate using miR156g reached 0.64%. For the target miRNA, the non-specific detection using miR156b was 5.5% while using miR156g-specific probes. These results demonstrate that the method has good generality and practicability for the detection of different miRNAs. The high specificity of ligation-based method depends on the complementary degree of the probes and the target sequence, but not the target miRNA sequence itself.

Pre-miRNA is the precursor of mature miRNA; it has stem-loop hairpin secondary structures that can be further cleaved by the dicer-like 1 (DCL1) and HYPONASTIC LEAVES1 (HYL1) into an miRNA: miRNA* (* represents secondary structure) duplex in the nucleus (Sun, 2012). To further investigate the pre-miRNA that were present in the sample on their specificity for miRNA detection, the pre-miRNA of miR156b (pre-miR156b) was prepared using the *in vitro* transcription reaction. The template concentration for the ligation reaction for pre-miR156b was determined to be in the range of 1 fmol–1 amol. The amplification plots of miR156b and pre-miR156b with the ligation-based qPCR assay showed that pre-miRNA produced negligible signals at concentrations below 10 amol (**Figure 5D**). In the template concentration range from 1 fmol to 10 amol, each C_T value of pre-miR156b was nearly ten cycles later than the corresponding miR156b, equating to 0.1% of the non-specific detection. These experimental data indicate that the developed ligation-based quantitative method has high specificity in distinguishing miRNAs from their corresponding pre-miRNAs.

Quantification of miR156b in Genetically Modified Soybean

As shown in **Figure 6A**, when soybean DNA was used as a template, no signal was produced in the reaction. This confirmed that the RNA sample reaction was not affected, even if some DNA was present. An equivalent amount of RNA extracted from GM and WT seedlings generated well-defined fluorescence signals in the ligation-based qPCR (**Figure 6B**), but it was evident that miR156b was expressed in GM soybean much more than in WT soybean. The synthesis miRNA156b range from 1 fmol to 0.0001 amol was selected for standard curve construction. As shown in **Figures 6C,D**, there was good linearity between C_T and the log of total RNA from 10 to 0.4 μg in the quantified assays. The data show that PCR efficiency was 101.7% and the R^2 of the standard curve was 0.979. Based on this standard curve, miR156b levels were determined in both the GM and WT samples, revealing that the fold change of miR156b expression level in the GM plants was 5.08 times higher using the established ligation-based quantitative method. These results are quite similar to the that of the developer's research data (4.82-fold change). Therefore, the established ligation-based method has the potential to detect and quantify GM samples using a miRNA mechanism.

CONCLUSION

In summary, this study experimentally demonstrated an ultrasensitive fluorescence method for miRNA quantification by elegantly integrating sensitivity and specificity. Under the hybridization procedure followed by DNA nick ligation catalyzed by T4 RNA ligase II, the initial template for miRNA amplification was obtained and quantified by quantitative real-time PCR. After optimizing the influential conditions, the proposed method could accurately discriminate among various miRNAs even when the difference was as small as a single nucleotide mismatch or the secondary structure difference leading by the nucleotides location site. High specificity was also achieved in the detection of mature miRNAs against their precursor sequences. In addition, the practicability of the proposed ligation-based quantitative PCR was further confirmed in that the total DNA and RNA samples from GM and WT soybeans generated well-defined fluorescence signals in a satisfactory dynamic range. The fold change of miR156b expression level compared well with data obtained by the conventional stem-loop qPCR, which further confirmed the reliability of the method. In contrast to the advantages and disadvantages of previously described methods for miRNA detection with data side-by-side (Table 2), the DNA probes and universal primers used in this method can be simply designed, without the need for complex structures such as the stem-loop reaction (Mei et al., 2012) or RCA assay (Liu et al., 2019), operators only need to substitute the specific sequences complementary to half of the target miRNA. Moreover, this method uses SYBR Green I and a small amount of ribonucleotide-modified DNA probes, meaning that it is also more cost efficient than that of the TaqMan probe. This characteristic greatly improves the method's efficiency and convenience in the quantification and detection of various miRNAs. Combining the detection limit and specificity results investigated with its ability to distinguish miRNA from the other similar sequences, the simple operation and feasibility of this ligation-based PCR

method implies that it can become a potential tool in GM content analysis.

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

YH contributed to the conception and design. LL and WY performed the main laboratory and wrote the manuscript. LD and WX carried out the data analysis and interpretation. CL provided the direction. FL drafted the manuscript and critically revised it for important intellectual content. All authors read and approved the manuscript.

FUNDING

Postdoc Foundation of Jilin Agricultural Technological Innovation Project, Jilin Academy of Agricultural Sciences, China (CXGC2020RCB002); High-level Talents Research Start-up Foundation of Jilin Agricultural Technological Innovation Project, Jilin Academy of Agricultural Sciences, China (CXGC2021RCG002); and the IAQST Foundation from Jilin Academy of Agricultural Sciences, Changchun, China (ZBS20Y02) supported this research.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.794752/full#supplementary-material>

REFERENCES

- Agrawal, A., Rajamani, V., Reddy, V. S., Mukherjee, S. K., and Bhatnagar, R. K. (2015). Transgenic plants over-expressing insect-specific microRNA acquire insecticidal activity against *Helicoverpa armigera*: an alternative to Bt-toxin technology. *Transgenic Res.* 24, 791–801. doi: 10.1007/s11248-015-9880-x
- Aslam, M., Sugita, K., Qin, Y., and Rahman, A. (2020). Aux/IAA14 regulates microRNA-mediated cold stress response in *Arabidopsis* roots. *Intern. J. Mol. Sci.* 21:8441. doi: 10.3390/ijms21228441
- Bally, J., Fishilevich, E., Doran, R. L., Lee, K., de Campos, S. B., German, M. A., et al. (2020). Plin-amiR, a pre-microRNA-based technology for controlling herbivorous insect pests. *Plant Biotechnol. J.* 18, 1925–1932. doi: 10.1111/pbi.13352
- Cao, D., Li, Y., Wang, J., Nan, H., Wang, Y., Lu, S., et al. (2015). GmmiR156b overexpression delays flowering time in soybean. *Plant Mol. Biol.* 89, 353–363. doi: 10.1007/s11103-015-0371-5
- Cheng, Y., Zhang, X., Li, Z., Jiao, X., Wang, Y., and Zhang, Y. (2009). Highly sensitive determination of microRNA using target-primed and branched rolling-circle amplification. *Angew. Chem. Intern. Edn.* 48, 3268–3272. doi: 10.1002/anie.200805665
- Chopperla, R., Mangrauthia, S. K., Rao, T. B., Balakrishnan, M., Balachandran, S. M., Prakasam, V., et al. (2020). A comprehensive analysis of MicroRNAs expressed in susceptible and resistant rice cultivars during *Rhizoctonia solani* AG1-IA infection causing sheath blight disease. *Intern. J. Mol. Sci.* 21:7974. doi: 10.3390/ijms21217974
- Curaba, J., Singh, M. B., and Bhalla, P. L. (2014). miRNAs in the crosstalk between phytohormone signalling pathways. *J. Exper. Bot.* 65, 1425–1438. doi: 10.1093/jxb/eru002
- De la Rosa, C., and Reyes, J. L. (2019). “Northern blot analysis of microRNAs and other small RNAs in plants,” in *Plant MicroRNAs. Methods in Molecular Biology*, Vol. 1932, ed. S. de Folter (New York, NY: Humana Press), 121–129. doi: 10.1007/978-1-4939-9042-9_9
- Dell'Aversana, C., Giorgio, C., and Altucci, L. (2017). MicroRNA expression profiling using agilent one-color microarray. *Methods Mol. Biol.* 1509, 169–183. doi: 10.1007/978-1-4939-6524-3_16
- Flores, T., Karpova, O., Su, X., Zeng, P., and Bilyeu, K. (2008). Silencing of GmFAD3 gene by siRNA leads to low α -linolenic acids (18:3) of fad3-mutant

- phenotype in soybean [*Glycine max* (Merr.)]. *Transgenic Res.* 17, 839–850. doi: 10.1007/s11248-008-9167-6
- Guo, C., Xu, Y., Shi, M., Lai, Y., Wu, X., Wang, H., et al. (2017). Repression of miR156 by miR159 regulates the timing of the juvenile-to-adult transition in *Arabidopsis*. *Plant Cell* 29, 1293–1304. doi: 10.1105/tpc.16.00975
- Jia, H., Li, Z., Liu, C., and Cheng, Y. (2010). Ultrasensitive detection of microRNAs by exponential isothermal amplification. *Angew. Chem. Intern. Edn.* 49, 5498–5501. doi: 10.1002/anie.201001375
- Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* 11, 597–610. doi: 10.1038/nrg2843
- Li, W., and Ruan, K. (2009). MicroRNA detection by microarray. *Anal. Bioanal. Chem.* 394, 1117–1124. doi: 10.1007/s00216-008-2570-2
- Liu, C., Han, J., Zhou, L., Zhang, J., and Du, J. (2020). DNazyme-based target-triggered rolling-circle amplification for high sensitivity detection of microRNAs. *Sensors* 20:2017. doi: 10.3390/s20072017
- Liu, X., Zou, M., Li, D., Yuan, R., and Xiang, Y. (2019). Hairpin/DNA ring ternary probes for highly sensitive detection and selective discrimination of microRNA among family members. *Anal. Chim. Acta* 1076, 138–143. doi: 10.1016/j.aca.2019.05.027
- Mei, Q., Li, X., Meng, Y., Wu, Z., Guo, M., Zhao, Y., et al. (2012). A facile and specific assay for quantifying MicroRNA by an optimized RT-qPCR approach. *PLoS One* 7:e46890. doi: 10.1371/journal.pone.0046890
- Miao, C., Wang, D., He, R., Liu, S., and Zhu, J.-K. (2020). Mutations in MIR396e and MIR396f increase grain size and modulate shoot architecture in rice. *Plant Biotechnol. J.* 18, 491–501. doi: 10.1111/pbi.13214
- Miao, C., Wang, Z., Zhang, L., Yao, J., Hua, K., Liu, X., et al. (2019). The grain yield modulator miR156 regulates seed dormancy through the gibberellin pathway in rice. *Nat. Commun.* 10:3822. doi: 10.1038/s41467-019-11830-5
- Papadopoulou, N., Devos, Y., Alvarez-Alfageme, F., Lanzoni, A., and Waigmann, E. (2020). Risk assessment considerations for genetically modified RNAi Plants: EFSA's activities and perspective. *Front. Plant Sci.* 11:445. doi: 10.3389/fpls.2020.00445
- Šečić, E., Kogel, K.-H., and Ladera-Carmona, M. J. (2021). Biotic stress-associated microRNA families in plants. *J. Plant Physiol.* 263:153451. doi: 10.1016/j.jplph.2021.153451
- Shu, P., Wu, C., Liu, W., Ruan, X., Liu, C., Hou, L., et al. (2019). The spatiotemporal expression pattern of microRNAs in the developing mouse nervous system. *J. Biol. Chem.* 294, 3444–3453. doi: 10.1074/jbc.RA118.004390
- Singroha, G., Sharma, P., and Sunkur, R. (2021). Current status of microRNA-mediated regulation of drought stress responses in cereals. *Physiol. Plant.* 172, 1808–1821. doi: 10.1111/ppl.13451
- Sun, G. (2012). MicroRNAs and their diverse functions in plants. *Plant Mol. Biol. Rep.* 80, 17–36. doi: 10.1007/s11103-011-9817-6
- Taverniers, I., Bockstaele, E. V., and Loose, M. D. (2004). Cloned plasmid DNA fragments as calibrators for controlling GMOs: different real-time duplex quantitative PCR methods. *Anal. Bioanal. Chem.* 378, 1198–1207. doi: 10.1007/s00216-003-2372-5
- Tian, W., Li, P., He, W., Liu, C., and Li, Z. (2019). Rolling circle extension-actuated loop-mediated isothermal amplification (RCA-LAMP) for ultrasensitive detection of microRNAs. *Biosens. Bioelectron.* 128, 17–22. doi: 10.1016/j.bios.2018.12.041
- Visentin, I., Pagliarini, C., Deva, E., Caracci, A., Turečková, V., Novák, O., et al. (2020). A novel strigolactone-miR156 module controls stomatal behaviour during drought recovery. *Plant Cell Environ.* 43, 1613–1624. doi: 10.1111/pce.13758
- Wang, B., Wang, J., Wang, C., Shen, W., Jia, H., Zhu, X., et al. (2016). Study on expression modes and cleavage role of miR156b/c/d and its target gene Vv-SPL9 during the whole growth stage of grapevine. *J. Heredity* 107, 626–634. doi: 10.1093/jhered/esw030
- Wei, S., Yu, B., Gruber, M. Y., Khachatourians, G. G., Hegedus, D. D., and Hannoufa, A. (2010). Enhanced seed carotenoid levels and branching in transgenic *Brassica napus* expressing the *Arabidopsis* miR156b gene. *J. Agric. Food Chem.* 58, 9572–9578. doi: 10.1021/jf102635f
- Yin, S., Ho, C. K., and Shuman, S. (2003). Structure-function analysis of T4 RNA ligase 2. *J. Biol. Chem.* 278, 17601–17608. doi: 10.1074/jbc.M300817200
- Zhang, J., Li, Z., Wang, H., Wang, Y., Jia, H., and Yan, J. (2011). Ultrasensitive quantification of mature microRNAs by real-time PCR based on ligation of a ribonucleotide-modified DNA probe. *Chem. Commun.* 47, 9465–9467. doi: 10.1039/c1cc13466c
- Zhang, P., Zhang, J., Wang, C., Liu, C., Wang, H., and Li, Z. (2013). Highly sensitive and specific multiplexed microRNA quantification using size-coded ligation chain reaction. *Anal. Chem.* 86, 1076–1082. doi: 10.1021/ac4026384
- Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y., and Robb, G. B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* 40:e54. doi: 10.1093/nar/gkr1263
- Zou, L., Li, R., Zhang, M., Luo, Y., Zhou, N., Wang, J., et al. (2017). A colorimetric sensing platform based upon recognizing hybridization chain reaction products with oligonucleotide modified gold nanoparticles through triplex formation. *Nanoscale* 9, 1986–1992. doi: 10.1039/c6nr09089c

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 © 2021 He, Long, Yan, Dong, Xia, Li 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.



WinRoots: A High-Throughput Cultivation and Phenotyping System for Plant Phenomics Studies Under Soil Stress

OPEN ACCESS

Edited by:

Xianzhong Feng,
Northeast Institute of Geography and
Agroecology, CAS, China

Reviewed by:

Aditya Pratap,
Indian Institute of Pulses Research
(ICAR), India
Pasquale De Vita,
Council for Agricultural
and Economics Research (CREA),
Italy
Ying Zhang,
Beijing Research Center
for Information Technology
in Agriculture, China

*Correspondence:

Shuo Li
lishuo@sdu.edu.cn

† These authors have contributed
equally to this work

Specialty section:

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

Received: 13 October 2021

Accepted: 19 November 2021

Published: 28 January 2022

Citation:

Zhang Y, Zhang W, Cao Q,
Zheng X, Yang J, Xue T, Sun W, Du X,
Wang L, Wang J, Zhao F, Xiang F and
Li S (2022) WinRoots:
A High-Throughput Cultivation
and Phenotyping System for Plant
Phenomics Studies Under Soil Stress.
Front. Plant Sci. 12:794020.
doi: 10.3389/fpls.2021.794020

Yangyang Zhang^{1†}, Wenjing Zhang^{1†}, Qicong Cao², Xiaojian Zheng¹, Jingting Yang¹,
Tong Xue¹, Wenhao Sun¹, Xinrui Du¹, Lili Wang¹, Jing Wang¹, Fengying Zhao¹,
Fengning Xiang¹ and Shuo Li^{1*}

¹ The Key Laboratory of Plant Development and Environmental Adaptation Biology, Ministry of Education, School of Life Sciences, Shandong University, Qingdao, China, ² Weifang Academy of Agriculture Sciences, Weifang, China

Soil stress, such as salinity, is a primary cause of global crop yield reduction. Existing crop phenotyping platforms cannot fully meet the specific needs of phenomics studies of plant response to soil stress in terms of throughput, environmental controllability, or root phenotypic acquisition. Here, we report the WinRoots, a low-cost and high-throughput plant soil cultivation and phenotyping system that can provide uniform, controlled soil stress conditions and accurately quantify the whole-plant phenome, including roots. Using soybean seedlings exposed to salt stress as an example, we demonstrate the uniformity and controllability of the soil environment in this system. A high-throughput multiple-phenotypic assay among 178 soybean cultivars reveals that the cotyledon character can serve as a non-destructive indicator of the whole-seedling salt tolerance. Our results demonstrate that WinRoots is an effective tool for high-throughput plant cultivation and soil stress phenomics studies.

Keywords: high-throughput plant cultivation, root phenotyping, whole-seedling phenotyping, plant stress phenotyping, abiotic stress

INTRODUCTION

Stress due to soil composition, such as salinity, heavy metals, and nutrient deficiency, is a primary cause for crop yield reduction (Miransari, 2010; McDonald et al., 2017; Zorb et al., 2019; Siao et al., 2020). This soil-related stress is first sensed by roots and affects their functions, leading to changes in plant physiology and biochemical reactions, ultimately influencing plant growth and yield (Li et al., 2018; Khan et al., 2019). The identification of genes linked to soil stress responses and assessment of how they contribute to combating stress drive molecular design breeding to develop stress-resistant and high-yielding crops, which have become vibrant topics for research. Crop soil stress tolerance is a complex trait that involves the regulation of a variety of genetic and non-genetic factors such as plant morphology, metabolism, and gene regulatory networks (Cui et al., 2015; Yang and Guo, 2018). However, fully revealing the underlying molecular genetic networks and effectively

harnessing them to guide molecular breeding has been extremely challenging. In recent years, the rapid development of high-throughput omics techniques, including phenomics platforms, has opened new avenues for the study of complex genetic traits in plants. Through integrated analyses such as phenotype-genotype association studies, it is now possible to comprehensively describe the plant genetic regulatory networks that respond to soil stress and apply this new knowledge to breed stress-resistant crops (Edwards and Batley, 2004; Grosskinsky et al., 2015; Jie, 2015; Al-Tamimi et al., 2016; Liu et al., 2016; Bevan et al., 2017; Muthamilarasan et al., 2019). At present, while next-generation sequencing technology has become more popular and accessible, high-throughput plant phenotyping has become the main bottleneck for large-scale functional analyses (Tardieu et al., 2017; Bolger et al., 2019; Wu et al., 2020).

Traditional crop phenotyping is generally carried out in the field, which is time-consuming and labor-intensive, suffers from low throughput, and can routinely be affected by natural environmental factors beyond the control of the researcher. Under such conditions, it is hard to obtain highly accurate phenotypic data to meet the requirements of phenomics studies. In past decades, several HTP (high-throughput phenotyping) platforms have been developed for use in the field or under controlled conditions. These platforms rely on automated transmission and control technologies. Various types of computer vision sensors, non-invasive imaging, and image analysis technologies can be used to perform phenotypic analysis on dozens, hundreds, or even thousands of plants over a short period of time (Furbank and Tester, 2011; Junker et al., 2014; Ghanem et al., 2015; Araus et al., 2018). The application of these tools to high-throughput phenotyping in controlled-environment facilities has the potential to improve accuracy and reduce the need for replication (Furbank and Tester, 2011). For example, the indoor automated robot phenotyping platform Phenoscope can accommodate the simultaneous cultivation of 735 *Arabidopsis* (*Arabidopsis thaliana*) plants and is equipped with a zenithal imaging system to monitor rosette size and expansion rate during the vegetative stage (Tisne et al., 2013). Similarly, the high-throughput rice (*Oryza sativa*) phenotyping identification facility (HRPF), which combines color imaging, X-ray computed tomography (CT), automatic controls, and image analysis functions, can extract at least 15 agronomic features from 533 rice cultivars grown in the greenhouse (Yang et al., 2014). PlantScreenTM (PSI, Czech Republic), a large-scale plant phenotyping imaging analysis platform equipped with a conveyor belt, was used to evaluate the changes in growth, morphology, and photosynthetic performance of nine *Arabidopsis* accessions exposed to salt stress (Awlia et al., 2016). These high-throughput phenotyping platforms can quickly obtain precise phenotypic values from potted plants, but their operation and maintenance costs are extremely high. In addition, research aimed at crop phenotyping usually focuses on the aboveground parts of plants while the acquisition of data from root morphology is much more limited (Chen et al., 2007; de Dorlodot et al., 2007; Meister et al., 2014; Wu et al., 2015; Crowell et al., 2016; Choudhury et al., 2019; Li M. et al., 2020; Nagel et al., 2020).

Roots, however, are the major route for water and nutrient uptake in plants, and they may also serve as storage organs for carbohydrates, as well as the direct sensing organ of soil stress (Zhu et al., 2011). Root morphology and root physiological functions change under soil stress (Ho et al., 2004; Zhu et al., 2005; Reynolds et al., 2007; York et al., 2013; Al-Tamimi et al., 2016). Therefore, phenotyping roots is an integral part of studying the plant phenome under soil stress conditions. To date, several root phenotyping technologies have been developed under controlled conditions. These can be divided into two primary categories. One subgroup relies on liquid growth medium (Ingram et al., 2012; Clark et al., 2013; Wu et al., 2018), germination paper (Le Marie et al., 2014; Atkinson et al., 2015; in 't Zandt et al., 2015; Gioia et al., 2017), solid growth medium (Busch et al., 2012), and other non-soil matrix growth systems. The second set of methods uses soil (Manschadi et al., 2006; Nagel et al., 2012). Non-soil matrix systems can easily achieve the direct quantification of root morphology and allow the non-invasive imaging analysis of numerous roots in a short time. However, non-soil technology cannot completely mimic the effects of soil on root development (Bengough et al., 2004; Hargreaves et al., 2009) and may even introduce unwanted variables such as hypoxia and bacterial growth (Zhu et al., 2011). To solve these problems, several groups have developed root phenotyping methods specifically designed for soil-grown plants. Although these methods achieved *in situ* observation and quantitative analysis of the root system by imaging transparent soil-filled columns or rhizotrons, their associated throughput was limited to fewer than 100 plants (Manschadi et al., 2006, 2008; Nagel et al., 2012). Advances in X-ray CT and magnetic resonance imaging (MRI) now allow non-invasive three-dimensional (3D) imaging of crop roots grown in opaque substrates that can quantify root system characteristics (Perret et al., 2007; Tracy et al., 2010; Clark et al., 2011; Mairhofer et al., 2013; Metzner et al., 2015; Pfeifer et al., 2015; Rogers et al., 2016; van Dusschoten et al., 2016). However, the high cost of CT and MRI equipment and the long scanning time needed for individual samples limit the application of these approaches in high-throughput root phenotyping platforms (Gregory et al., 2009; Zhu et al., 2011). Therefore, developing a high-throughput soil cultivation and whole-plant phenotyping system suitable for crop soil stress research remains a largely unfinished task.

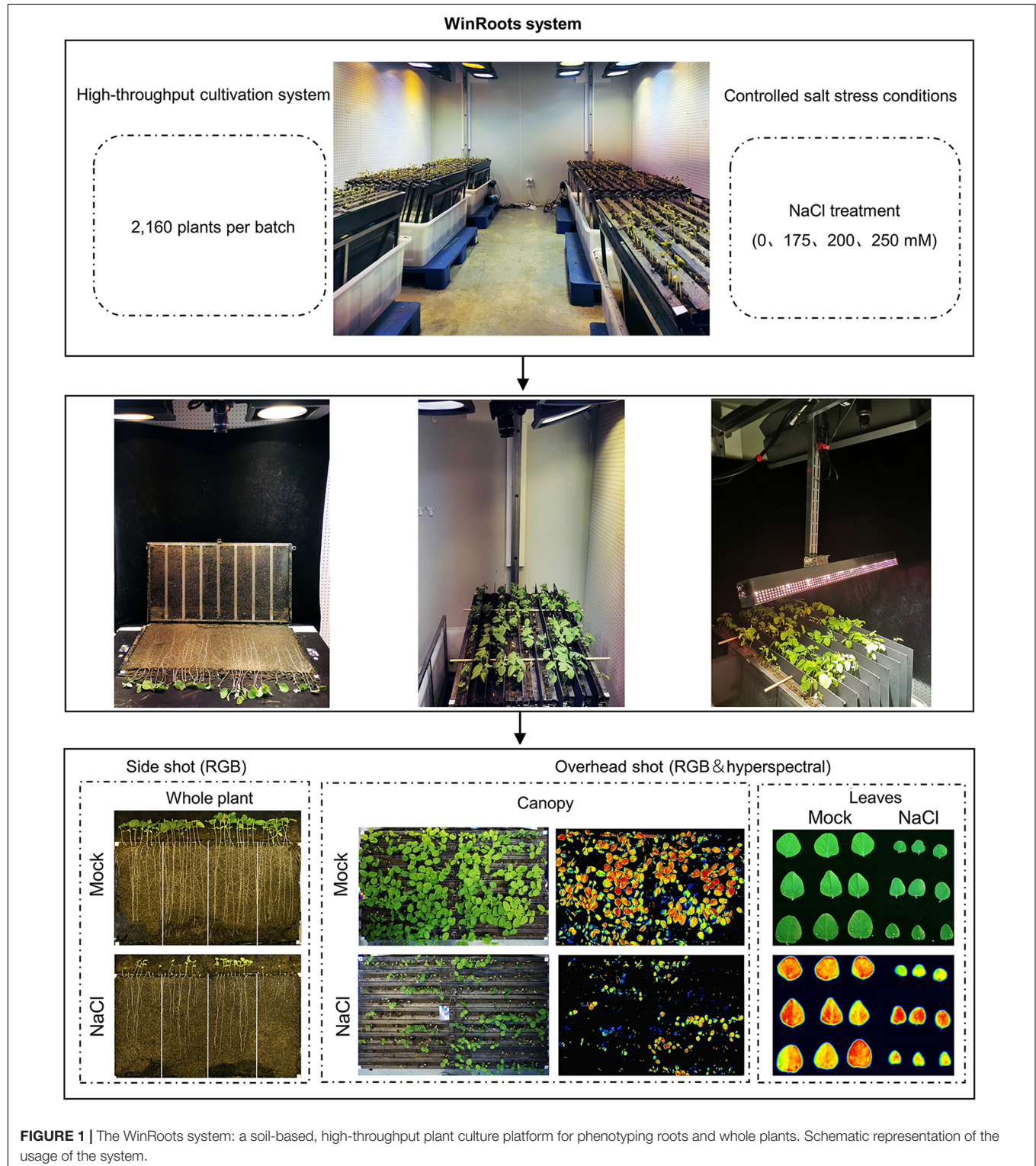
Here, we describe the recently developed WinRoots platform, a high-throughput plant cultivation and phenotyping system for plants grown in a soil matrix. WinRoots allows the dynamic monitoring and acquisition of whole-plant phenotypes from plants grown in high throughput in uniform and controllable soil stress conditions. It represents an economical and flexible system with significant space savings compared to other methods. We phenotyped soybean (*Glycine max*) plants exposed to salt stress as a proof of concept and demonstrated the uniformity and controllability of soil stress conditions and the high throughput of the WinRoots system. We developed optimized salt stress conditions and phenotypic indices suitable for high-throughput phenotyping of salt tolerance in soybean. In addition, a high-throughput

multiple-phenotypic analysis indicates that the cotyledon character can serve as a useful non-destructive indicator of the whole-seedling salt tolerance in soybean. This new system will be a powerful tool for studying the complex traits of crops in response to soil stress.

MATERIALS AND EQUIPMENTS

Structure of WinRoots System

As shown in **Figure 1**, the WinRoots system consists of two or more independent circulatory systems and several plant



culture units. The circulatory system is composed of electric pumps, an ultraviolet (UV) sterilizer, and pipes equipped with ball valves (**Figure 2A**). When turned on, the electric pumps promote the circulation of the liquid in the system when in use. It also help discharge the liquid out of the system in between experiments. A UV sterilizer irradiates the liquid flowing through the water pipes to inhibit the growth of algae, bacteria, and other microorganisms, thus avoiding contamination of the liquid environment that would otherwise affect root growth. The pipes are fitted with ball valves and connect the culture boxes; opening or closing the ball valve can form a shared and homogeneous liquid environment between boxes or an independent liquid environment for each culture box (**Figure 2A**). In this experiment, the circulatory system worked 1 h per day every 12 h to maintain the uniform of the electrical conductivity in the soil matrix.

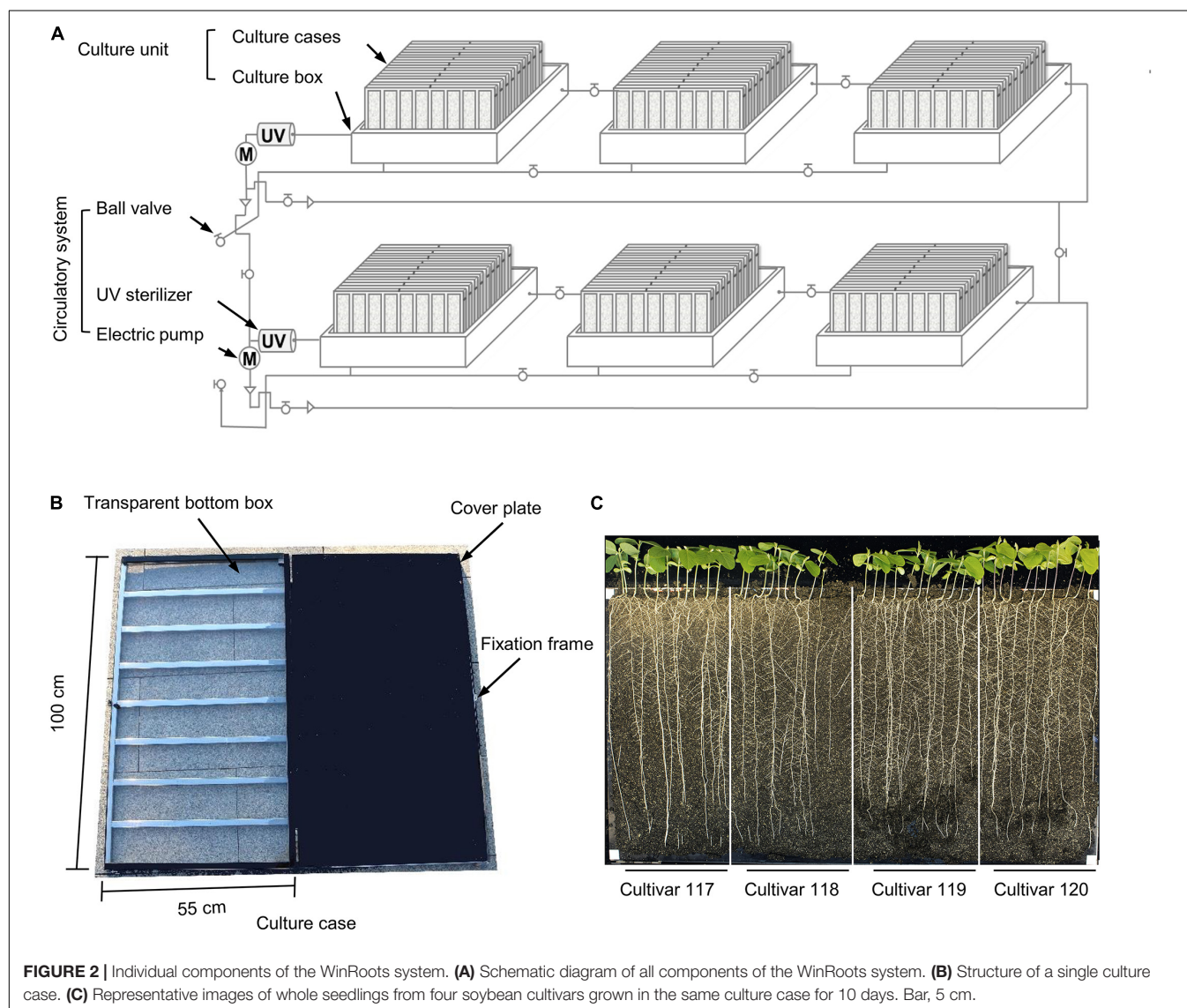
Each culture unit comprises several culture cases stacked obliquely in a culture box. The inner wall of each culture box is

106 cm (length) \times 79 cm (width) \times 32 cm (height). Each box can accommodate nine culture cases stacked obliquely. Each culture case consists of a 100 cm \times 55 cm \times 2 cm box with a transparent bottom and a cover plate of the same size (**Figure 2B**). There are 120 vent holes with a 1-mm diameter, evenly distributed over the surface of the cover plate. To maintain a constant distance of 2 cm between two consecutive plates inside the culture case, a metal fixation frame surrounds each culture case (**Figure 2B**).

The transparent bottom box is filled with soil matrix and is then assembled by placing the cover plate over the transparent box and fixed with an openable clip frame to provide the soil environment for plant growth. The culture box serves as a container for water or culture solution, from which the soil in the culture cases can draw water through capillary action.

Plant Material

Experiments were conducted on plants from 178 soybean cultivars harvested in 2019. The seeds were surface sterilized



in 75% ethanol for 3 min, rinsed in sterile water three times, and then placed in a moist seed germination box to germinate at 25°C in the dark. After 48 h, seedlings with visible radicle elongation of about 1 cm were selected and transferred to culture cases and planted at a depth of 2 cm. The matrix in the culture cases was composed of nutritive soil and vermiculite at a ratio of 3:1, mixed with a fixed volume of water or NaCl solution at a specific concentration. The WinRoots system was placed in the climate chamber. Environment parameter settings were as follows: light/dark = 16 h/8 h, temperature 25°C, and humidity 75%. Though the system can support the soybean seedlings grow normally to stage V₂, in this experiment, the seedlings grow to stage V₁ (~10 days after planting).

METHODS

Experimental Design

The version of the WinRoot system presented here is equipped with six culture units, each of which contains nine culture cases that allow visual inspection of root growth. Each culture case can accommodate 40 soybean seedlings for normal growth (with spacing between neighboring seedlings of about 2.5 cm), and each experiment can accommodate up to 2,160 soybean seedlings for simultaneous growth (Figures 1, 2A,C and Supplementary Figure 1). By adjusting the number or size of culture units and culture cases, the throughput of the WinRoots system can be adjusted at the user's discretion to suit their space and plant material requirements. When setting up an experiment, the emerging radicle of a soybean seed should be placed close to one side of the transparent bottom box. The planting design for uniformity assay and salt stress condition determination was shown in Supplementary Figure 1.

The culture case should be positioned obliquely so that the transparent side forms a 66° angle with the ground and remains in this position during the entire course of the experiment. Root geotropism will promote growth along the transparent side, thus allowing the unobstructed observation of plant roots.

Before the start of the experiment, it is critical to ensure that the liquid level in all culture boxes reaches 10 cm above the bottom of the culture cases; six culture boxes are placed symmetrically in pairs and arranged on the floor support of the climate chamber at equal intervals. The culture boxes located on the same side are connected to each other through the water inlet and outlet pipes with ball valves. Opening ball valves allow liquid circulation across adjacent culture boxes and ensure that the soil in all visible culture boxes receives the same uniform and stable external liquid environment.

Each culture case in the culture box was divided into four site intervals. Each site comprised 10 seedlings from one cultivar. The primary roots of soybean seedlings cultured in the WinRoots system grew straight down along the transparent wall. After 10 days of growth under mock conditions, the root tips reached the bottom of the culture case. Red-green-blue (RGB) images or hyperspectral data can be collected for whole seedlings from the transparent side of the case or for the seedling canopy from the top of the culture units during

cultivation, respectively (Figures 1, 2C), and then analyzed to extract phenotypic trait values.

Conductivity Tests

A portable handheld soil conductivity meter (0–10,000 μs/cm) was used to measure soil conductivity. Soil conductivity was measured about 5 cm below the soil surface at four sites 25 cm apart in each culture case, at 1 day before planting and 10 days after planting.

Red-Green-Blue and Hyperspectral Data Acquisition

A Canon EOS 700D digital camera and a Resonon PIKA L hyperspectral imager were used to collect RGB (in raw format) and hyperspectral images, respectively. The camera was positioned from a slidable horizontal rail 1.5 m above the plant material. The images of soybean canopy (overhead shot) and of the whole seedlings (side shot) were collected every day. On the 9th day of cultivation, images were taken of detached leaves with three replicates for each cultivar.

Phenotype Measurement

ImageJ version 1.53a¹ software was used to measure primary root length (PRL) and leaf area from RGB images (in tiff format). After imaging whole seedlings on the 9th day, the shoot part and cotyledons of each seedling were harvested separately and their fresh weight determined. The samples were then dried at 100°C for 3 days and their dry weight measured. We calculated the water content and water weight per plant based on the formula $WC = (\text{average fresh weight} - \text{average dry weight}) / \text{average fresh weight}$ and $WW = \text{average fresh weight} - \text{average dry weight}$. The relative index = the index of seedlings grown under salt stress/the index of seedlings grown in mock condition.

Data Analysis

Microsoft Excel 2016 with XLSAT plug-in (Ver. 2019.2.2) was used for data analyses. The Student's *t*-test was used to check whether two sets of data differ significantly. For correlation analysis, the Spearman correlation test was performed on the phenotypic data.

For agglomerative hierarchical clustering, the phenotypic indices were standardized, the similarity was evaluated with Kendall correlation coefficient, and the unweighted pair-group average agglomeration method was used.

RESULTS

Uniform Soil Cultivation Environment in WinRoots System

Most high-throughput crop phenotyping efforts are performed in pots (one plant per pot) or on germination paper (several plants per sheet of paper) (Adu et al., 2014; van Dusschoten et al., 2016), making it difficult to ensure that the “micro-environment”

¹<https://imagej.net/>

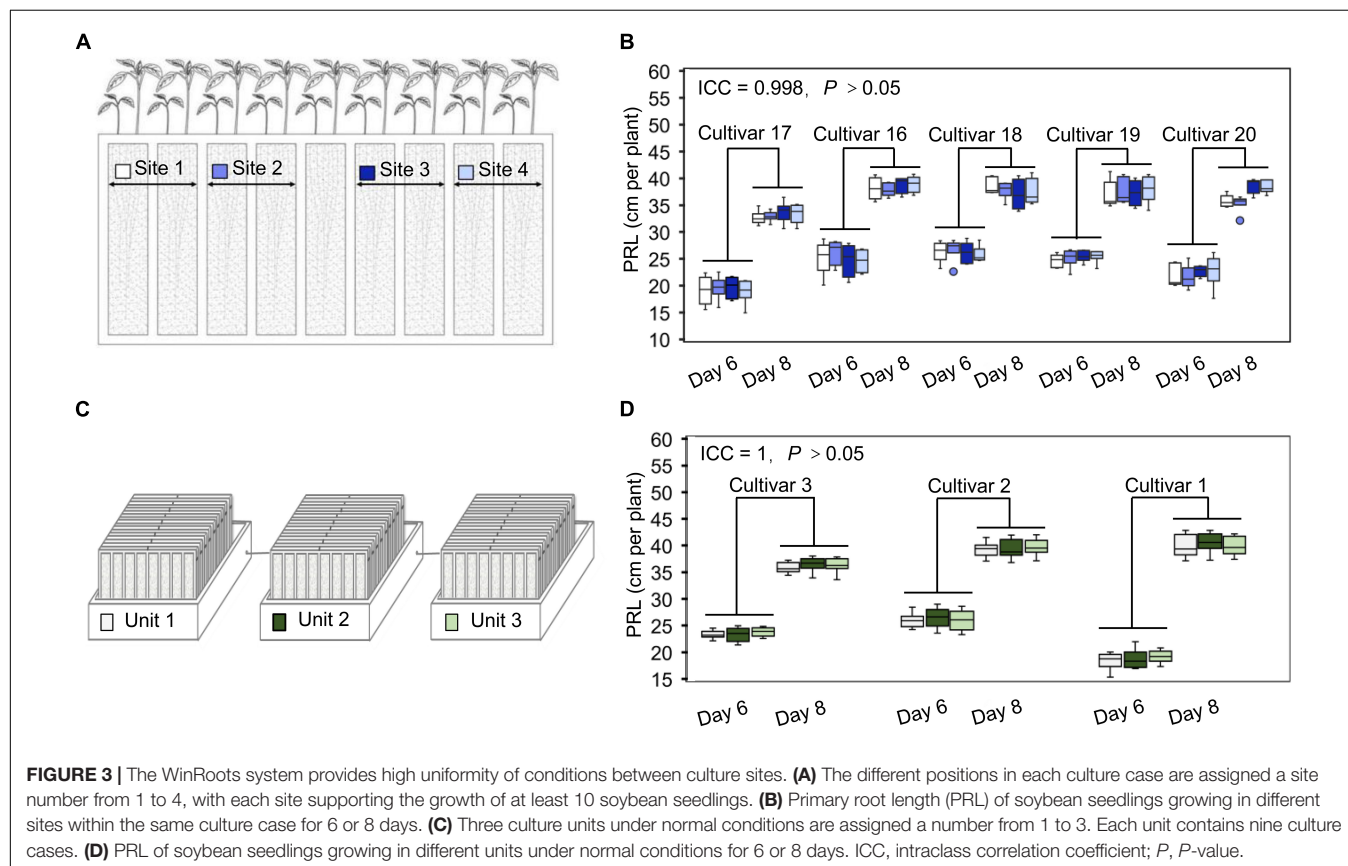
of these materials is consistent throughout the entire system. To solve this problem, we established WinRoots (**Figures 1, 2A**), a high-throughput plant cultivation and phenotyping system that is based on a soil matrix and offers root visibility.

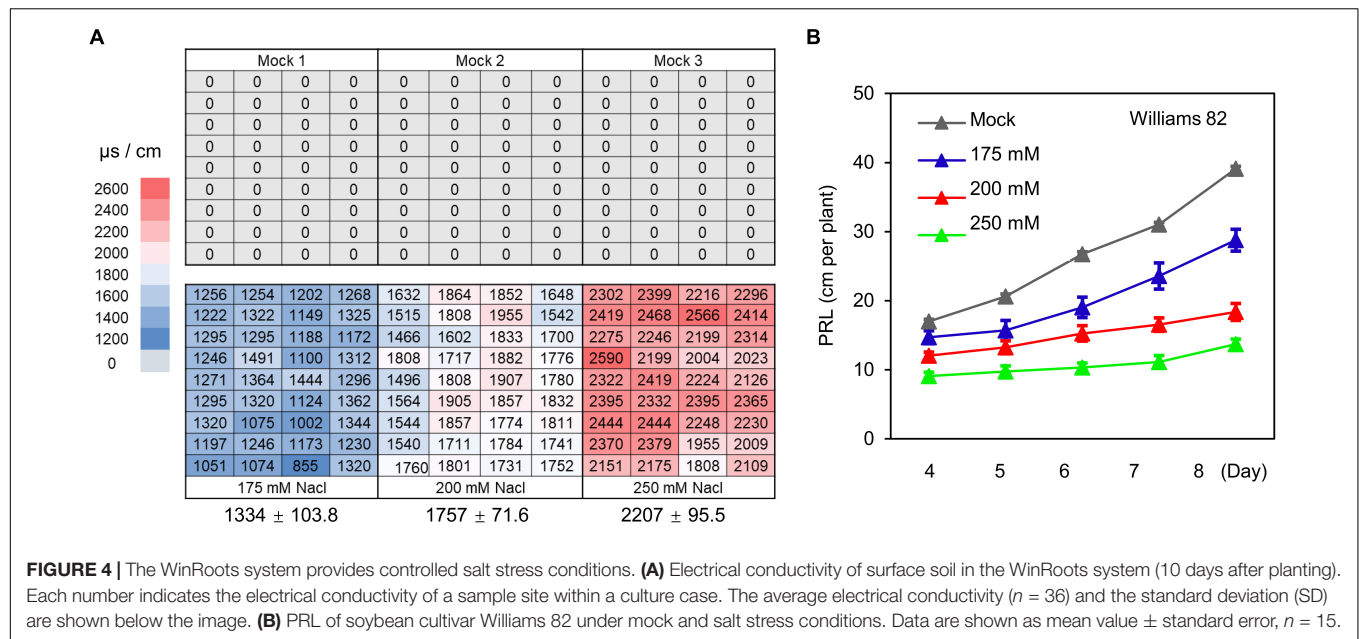
Changes in environmental factors can drastically affect measured plant phenotypes. To partially account for these effects and accurately identify phenotypic differences between cultivars, one strategy is to include a large number of replicates to dilute any variation caused by an uneven environment. However, this will also increase the cost of phenotyping. An alternative way of obtaining accurate phenotypic data at an affordable cost is the establishment of highly controlled and stable cultivation conditions with limited repetitions. To evaluate the uniformity of the soil cultivation environment in the WinRoots system, we used soybean PRL as an indicator to analyze the potential for position effects at different sites within one culture case and across different culture boxes. For this test, the WinRoots system was placed inside a climate-controlled growth chamber to minimize variation in temperature and humidity and used to collect data for 520 seedlings from 7 soybean cultivars. We determined PRL for four soybean cultivars placed at four different sites after 6 and 8 days of growth to test the position effect within culture cases, each site containing 10 replicate seedlings. Our results showed that root length was very consistent for each cultivar across the four sites tested, as the intraclass correlation coefficient (ICC) was 0.998, with an associated P -value of > 0.05 (**Figures 3A,B** and **Supplementary Figure 1**). Similarly, we evaluated PRL for

three soybean cultivars across three different culture boxes to assess the possible contribution of the culture boxes to variation. Again, root length displayed high consistency, with an ICC of 1 and $P > 0.05$ (**Figures 3C,D**). These results indicate that the WinRoots system can provide a uniform soil culture environment for soybean seedlings and that the phenotypic data obtained under these conditions are accurate and reliable.

Precise Control of Soil Salt Stress Conditions

Soil stress conditions in the field are often uneven, which constitutes a major limiting factor for obtaining accurate phenotypes related to soil stress in field-grown crops. To test whether the WinRoots system can provide soil salinity conditions that are precisely controlled and highly homogeneous, we evaluated soil conductivity at three different salt stress levels. We used an equal volume of water (mock) or NaCl solution (175, 200, and 250 mM) to soak the nutritive soil-vermiculite mixture and then fill the culture cases, and then placed the culture cases in different culture boxes (**Supplementary Figure 1**). Germinating soybean seeds were sown after the circulatory system had been allowed to run for 1 day. We planted seedlings at the same positions within each unit for mock and salt stress conditions (**Supplementary Figure 1**). Soil conductivity was measured for a total of 216 sites sampled, 1 day before and 10 days after planting (**Figure 4A** and **Supplementary Figure 2**). Although





we observed significant differences in soil conductivity between different conditions, as expected, the variation in conductivity at different positions for a given condition was very small (**Figure 4A** and **Supplementary Figure 2**). This indicates that soil conditions in the WinRoots system are uniform and controllable when the system is used to impose high salinity stress.

Next, we determined PRL for the soybean cultivar Williams 82 exposed to the above salt stress conditions for 4–8 days. We observed an inhibition of primary root growth by all salt stress conditions that was commensurate with the salt concentration. The difference in PRL between seedlings exposed to various salt concentrations increased gradually over time (**Figure 4B**). Together, these results indicate that the WinRoots system can evoke uniform and controlled salt stress when seedlings are grown in a soil matrix and allows the collection of precise seedling phenotypes.

Determination the Optimal Salt Stress Condition in Soybean Salt Tolerance Screen

In a larger-scale experiment, we used root length, fresh weight, and water content of 11 soybean cultivars under three salt stress conditions in the WinRoots system to define the appropriate conditions to phenotype salt stress tolerance.

The structure of the root system at the seedling stage largely reflects the growth of newly emerged seedlings experiencing stress to support seedling growth and development (de Dorlodot et al., 2007; Meister et al., 2014; Wu et al., 2015). To evaluate the differences in soybean seedlings under different salinity levels, we measured the PRL of 720 individual seedlings from 11 soybean cultivars either grown under normal conditions (mock treatment) or exposed to 175, 200, or 250 mM NaCl for 10 days. To evaluate the salt stress sensitivity of the seedlings, we use the relative root length (RRL) as an index. Soybean PRL was

significantly lower in all three salt stress conditions tested here than in the mock treatment (**Figure 5A**, $P < 0.001$). In addition, PRL showed greater variation at the salt concentrations of 175 and 200 mM NaCl, suggesting that these two conditions may be used as sensitized stress conditions to screen root-related salt stress phenotypes with the WinRoots system. Importantly, all recorded data followed a normal distribution in all conditions (**Figure 5B**, $P > 0.05$).

We also measured the fresh weight (FW), the water content (WC) and calculated their relative indices (relative fresh weight, RFW; relative water content, RWC) of the aboveground parts of soybean seedlings. As with PRL, all salt stress conditions significantly reduced FW and WC relative to mock conditions, and at 175 mM NaCl condition the indices showed higher variation (**Supplementary Figures 3A, 4A**, $P < 0.001$). All indices followed a normal distribution (**Supplementary Figures 3B, 4B**, $P > 0.05$).

Based on the above observations, we determined that 175 mM NaCl was a suitable salt stress condition to quantify salt tolerance and sensitivity in a soybean population by the WinRoots system.

High-Throughput Multiple-Phenotypic Assay Reveals Cotyledon Character as a Useful Indicator for Salt Tolerance Estimation in Soybean

To access the soybean salt tolerance on a multiple-phenotype scale, we turned next to the high-throughput analysis of 146 soybean cultivars and recorded whole-seedling phenotypes such as the side (on day 10th) and leaf images (on day 9th) after planting (**Supplementary Figure 5**). We measured the PRL and leaf area (LA) of plants in both mock and NaCl treatment. In addition, we measured the FW and DW of the cotyledons and shoots for the same cultivar. Based on these data, we obtained 18

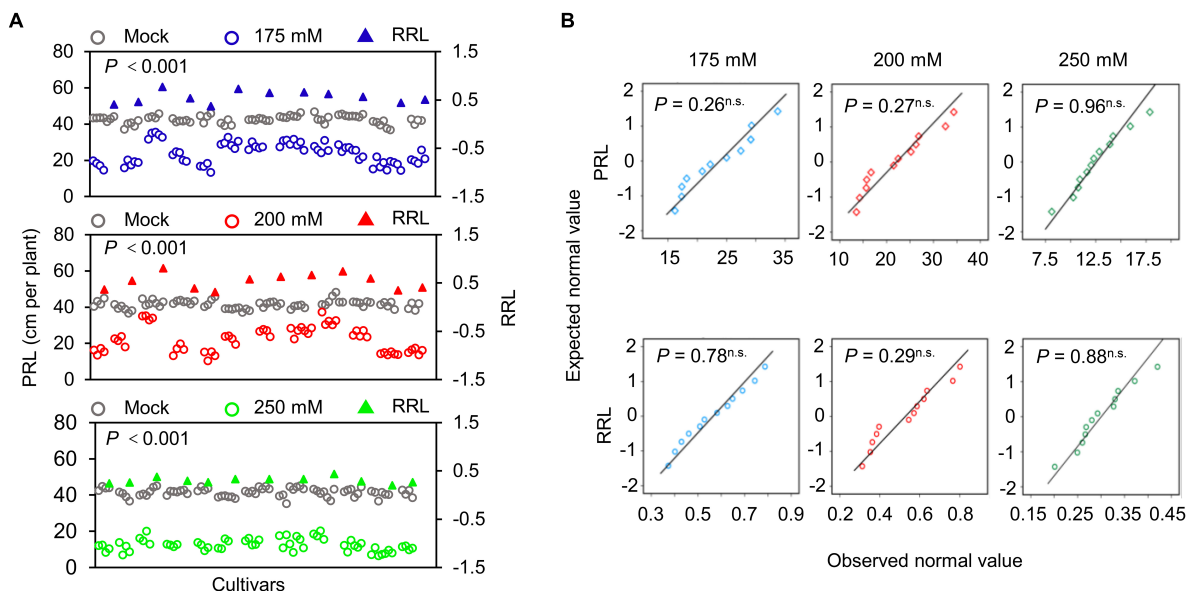


FIGURE 5 | Determination of primary root length. **(A)** Scatter distribution map of PRL (circles) and relative primary root length (RRL, triangles) of soybean seedlings. Gray, blue, red, and green represent the soybean seedlings grown under mock treatment or NaCl treatment of 175, 200, or 250 mM, respectively. Each circle represents the PRL of an individual seedling. Each triangle represents the value of the RPL of a soybean cultivar. Data were collected 10 days after planting. A paired Student's *t*-test was conducted between mock and NaCl-treated seedlings to test for significance. **(B)** Quantile-quantile plots of PRL and RRL. The normality of the data was tested using the Shapiro-Wilk test. *P*, *P*-value; n.s., not significant. The planting design is shown in **Supplementary Figure 1**.

indices to access the state of soybean seedlings in mock condition and under salt stress. The indices included the PRL, LA, FW of shoots or cotyledons, water weight (WW) of shoots or cotyledons, and their corresponding relative indices (salt stress/mock). All these indices were salt stress affected (**Supplementary Figure 6**).

To find out the relationship among these phenotypic indices, we performed a correlation analysis. We observed significant positive correlations between the phenotypic indices of plants grown under the same conditions while the correlation coefficients varied (**Figure 6A**). This implies that the phenotypic indices are related to each other and are different. Notably, the cotyledon-related indices showed very high correlation coefficients with some typical salt tolerance-related indices. For example, the WW of cotyledons under salt stress (NWW_C) showed correlation coefficients of 0.90 and 0.74 with the PRL and FW of shoots under salt stress (NRL and NFW_S), respectively (**Figure 6A**). It showed similar results with the relative WW of cotyledons (RWW_C). To explore whether the NWW_C or RWW_C can be used to predict the other salt tolerance indices, we constructed the one-dimensional linear regression equations (**Supplementary Figure 7**) and used them to predict the NFW_S, RFW_S, NRL, and RRL in a new sample group which included 32 cultivars. As a result, the predicted NFW_S showed a coefficient of determination up to 0.85 with its observed value. For RFW_S, NRL, and RRL, the coefficient of determination was 0.69, 0.61, and 0.51, respectively (**Figure 6B**). These results showed that the moisture in cotyledons could be used to predict the shoot- or root-related indices in soybean salt tolerance screens.

For dry seeds, the seed weight (SW) approximately equals the amount of dry material in cotyledons. In correlation analysis, the

SW showed a significant negative correlation with the relative FW in shoots (RFW_S) and the relative WW in shoots (RWW_S) (**Figure 6A**), which suggested that the dry material in cotyledons also correlate the salt tolerance of plants. We further performed agglomerative hierarchical clustering with the dry-weight-related indices of cotyledons and other salt-tolerance-related indices among the 146 cultivars. As a consequence, the population was grouped into two major clusters (**Figure 7A** and **Supplementary Figure 8**). The cultivars in cluster 1 shared common characters such as heavier cotyledons (before germination and after the NaCl or mock treatment), higher salt tolerance, and lower salt sensitivity relative to the cultivars in cluster 2 (**Figures 7B,C**). As taking off a single cotyledon from a young seedling would not affect its viability, these results demonstrated that the character of cotyledons could serve as a non-destructive indicator for salt tolerance estimation in soybean.

In conclusion, the WinRoots system provides uniform and controllable soil stress conditions for seedling growth, can be used for high-throughput cultivation and phenotyping under soil stress, and helps provide accurate and diverse soil stress-related phenotypic data. WinRoots therefore offers an improved method to analyzing complex traits such as soil stress.

DISCUSSION

The collection of high-quality data in a high-throughput manner is essential for the study of complex traits. Here, we described WinRoots, a high-throughput crop seedling cultivation and phenotyping system based on a soil matrix, and its

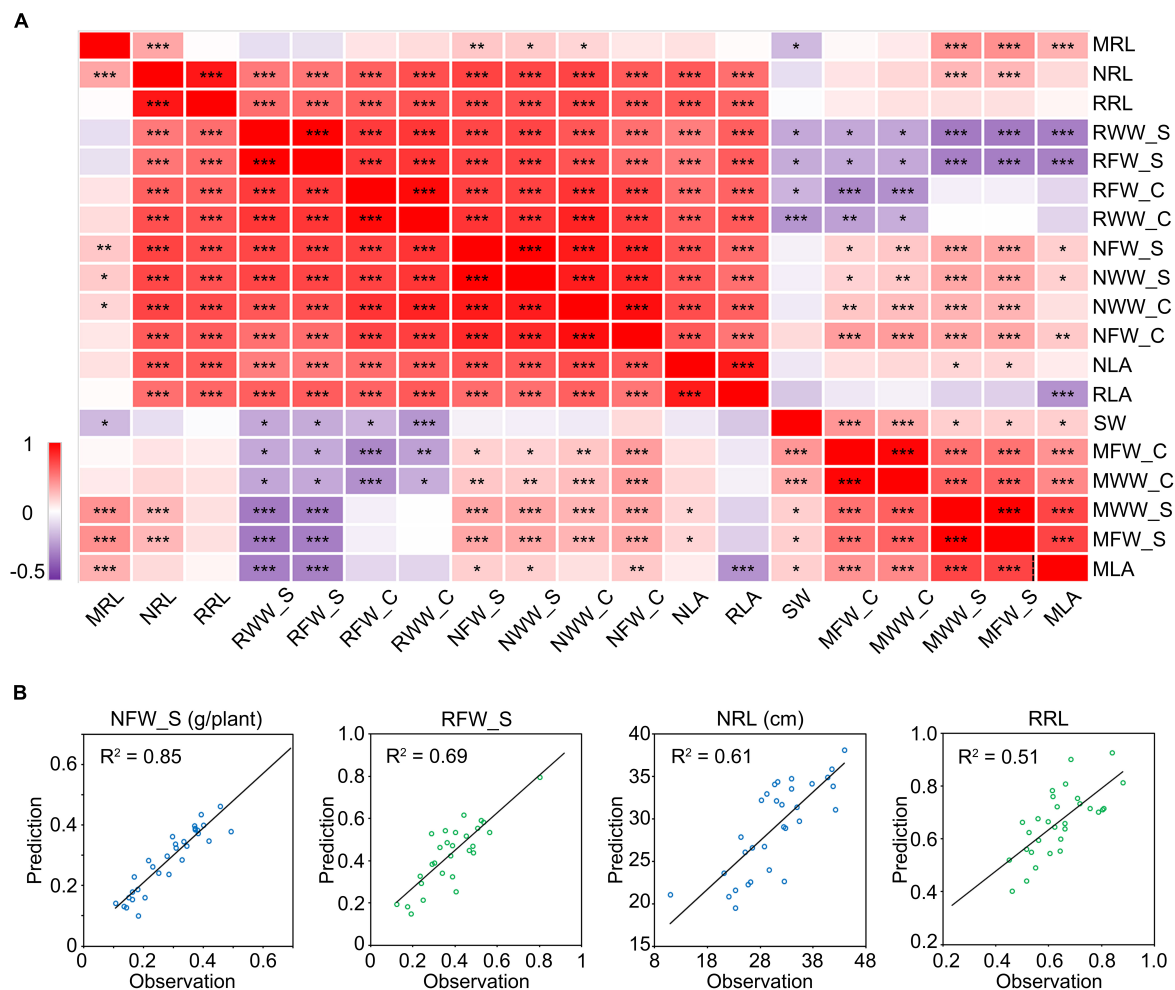


FIGURE 6 | Correlation analysis between salt stress-related traits. **(A)** Correlation matrix between salt stress-related traits. The numbers in the squares indicate the Pearson correlation coefficients. Data were collected from 10 seedlings of 146 different cultivars grown for 10 days under mock or 175 mM NaCl treatment. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. **(B)** Regression curve between the predicted value and observed value. The prediction was performed in a new group including 32 cultivars. The first letter “M” in the index name means “mock treatment,” “N” means “NaCl treatment” and “R” means “relative index” (NaCl treatment/mock treatment). The suffix “_C” represents “cotyledon” and “_S” represents “shoot.” RL, root length. LA, leaf area. FW, fresh weight. WW, water weight. SW, seed weight.

application to whole-seedling phenotyping. Capitalizing on the highly controlled and homogeneous soil salt stress environment provided by WinRoots, we collected and analyzed a cultivar of image- and physiological-based traits of soybean seedlings and demonstrated its application potential for high-throughput phenomics studies under soil stress conditions.

WinRoots System Enables the Accurate Quantification of the Whole-Seedling Phenome

Prerequisites to achieving an accurate quantification of the plant phenome are high-throughput plant cultivation and fast and accurate phenotype collection. In the case of the phenotyping platform GROWSCREEN-Rhizo, 72 individual plants can be grown at once, with a phenotyping rate of 60 plants per hour (Nagel et al., 2012). The high-throughput rice phenotyping

facility (HRPF) was reported to have the capacity to hold 5,472 pots, while its associated rice automatic phenotyping platform (RAP) can collect at least 15 agronomic traits at the speed of 80 pots each hour (Yang et al., 2014). The WinRoots system can accommodate 2,160 soybean seedlings (about 12-day-old at normal condition) simultaneously in a 12-m² climate chamber, achieving a growth density of 180 soybean plants per square meter. With the use of an RGB digital camera, the WinRoots system can acquire images for at least 800 seedlings every hour, while the hyperspectral camera can collect 720 canopy images or images from 81 detached leaves *in vitro* per hour with the human factor as the limitation (Figure 1). Thus, the WinRoots system allows (1) high-throughput plant cultivation in soil using only a small footprint and (2) fast acquisition of digital traits from whole seedlings.

Obtaining high-quality phenotypic data is both a prerequisite and a very challenging aspect of performing phenotype-genotype

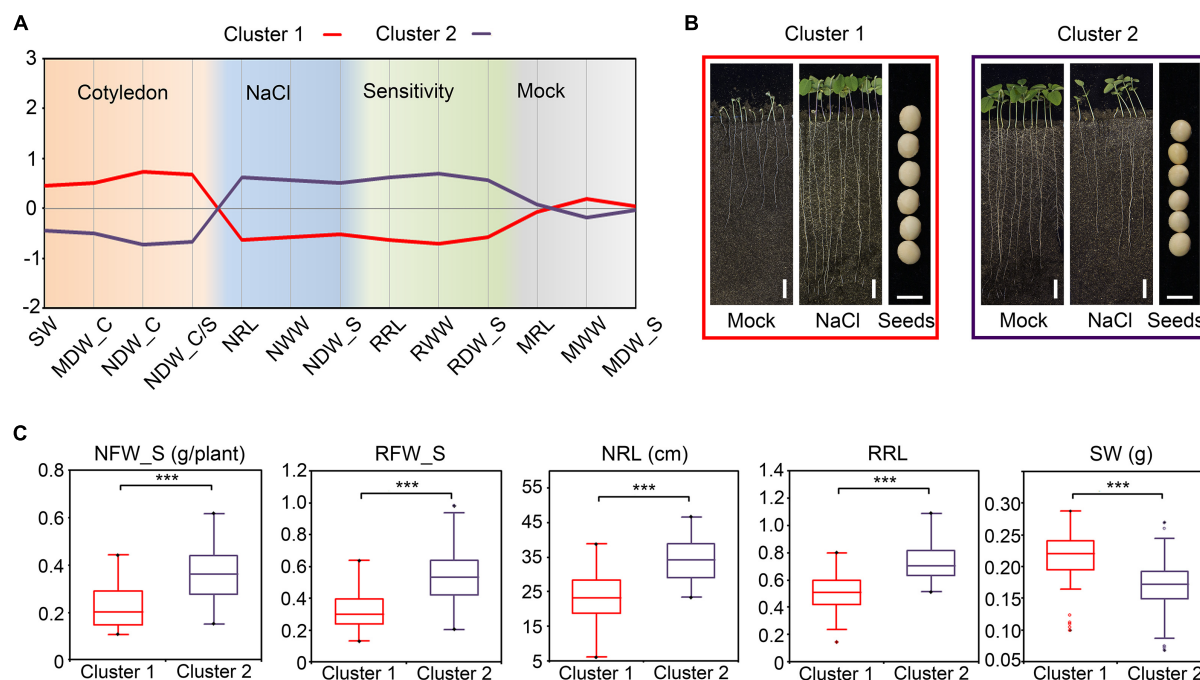


FIGURE 7 | Agglomerative hierarchical clustering of salt stress-related traits in soybean. **(A)** The profile plot of the agglomerative hierarchical clustering of salt stress-related traits in soybean. Analysis was performed with the standardized indices. **(B)** Phenotypes of the representative cultivar for Cluster 1 and Cluster 2. Images of the whole seedlings from the soybean cultivars grown in WinRoots for 10 days. Bar, 5 cm in the whole-seedling images. Bar, 5 mm in the seeds images. **(C)** The comparison between the indices of Cluster 1 and Cluster 2. *** $P < 0.0001$ ($n = 146$), un-paired Student's t -test.

association analyses to identify loci underlying quantitative traits. For rice and cotton (*Gossypium hirsutum*), high-quality phenotypic data obtained with high-throughput phenotyping platforms were subjected to genome-wide association analysis (GWAS), which not only identified the positions of known traditional quantitative traits, but also revealed many new loci, highlighting the strong advantages of this method (Yang et al., 2014; Al-Tamimi et al., 2016; Li B. et al., 2020). Here, the high-throughput multi-phenotypic analysis revealed that cotyledon-related indices as indicators for whole-seedling soybean salt tolerance estimation. We precisely predicted the other salt tolerance-related indices using the cotyledon-related indices (Figure 6). This indicates that the phenotypic indices obtained with the WinRoots system can tease apart the standing variation between soybean cultivars, thereby providing high-quality phenotypic data for later GWAS approaches.

Taken together, the WinRoots system is a powerful tool for the accurate quantification of phenotypes in a high-throughput manner.

Application of the WinRoots System to Crop Phenomics Studies Under Soil Stress

The development of phenotyping systems that can provide stable and uniform soil stress conditions is one of the main technical challenges in high-throughput crop phenomics under soil stress conditions. Currently, most crop phenotyping studies

assessing the effects of soil stress are conducted in the field. However, the uneven distribution of salinity, heavy metal ion concentrations, water content, and other environmental factors in the field severely hinder the evaluation of stress tolerance in crops. It therefore is often necessary to collect phenotypic and environmental data across multiple locations and seasons with high replication, which greatly increases the research cost (Luo et al., 2017; Tardieu et al., 2017; Morton et al., 2019). In crop phenomics studies, several methods have been developed to limit the impact of environmental heterogeneity on plant phenotypes. For example, the Scanalyzer 3D phenotyping platform maintains a constant water moisture level in the soil throughout cultivation with an automated weighing and watering system that adjusts irrigation to compensate for evaporation (Al-Tamimi et al., 2016). The Phenoscope system avoids environmental heterogeneity by rotating pots continuously, resulting in an improved reproducibility for phenotypic data (Tisne et al., 2013). In contrast to these platforms, the WinRoots system uses conductivity monitoring and a liquid circulation system to achieve uniformity of soil stress conditions within and between cultivation units. Soil conductivity under three salt stress conditions was well controlled and maintained as uniform in the WinRoots system (Figure 4A and Supplementary Figure 2). Additionally, the WinRoots system is very flexible in terms of applied stress conditions, as multiple stressors may be imposed on seedlings to meet experimental requirements (for example, we exposed seedlings to four levels of salt stress conditions in this study). Similarly, the number of cultivation units and

liquid circulation systems can be configured to accommodate the specifications of the greenhouse and climate chamber, so as to meet the needs of diversified crop soil stress research.

The effects of soil stress on phenotypes are multi-level and multi-faceted. However, due to the lack of powerful phenotypic tools, most past studies on crops experiencing soil stress have been based on a single phenotype (Mano and Takeda, 1997; Campbell et al., 2015; Awlia et al., 2016; Luo et al., 2017). In recent years, with the emergence of high-throughput phenotyping platforms, multi-phenotype studies have offered a number of advantages for elucidating complex traits (Yang et al., 2014; Al-Tamimi et al., 2016; Li B. et al., 2020). For example, Li B. et al. (2020) used an automated phenotyping platform to collect 119 image-based digital traits (i-traits) on natural populations of cotton grown under drought stress conditions and successfully identified 390 drought-related quantitative trait loci (QTLs) by GWAS (Li B. et al., 2020). Our study established that several phenotypic indices were highly correlated in soybean seedlings under salt stress, including root length, fresh weight, and water content, while others were not, such as leaf area and emergence rate. Our results further indicated that phenotypes can be correlated or independent, reflecting the complexity of the phenotypes associated with salt tolerance (**Supplementary Figure 9**). Therefore, it is critical to collect multiple phenotypes rather than a single one for the study of complex traits. The WinRoots system can collect all phenotypic data based on RGB images, hyperspectral data, and plant materials. For instance, data for root length, plant height, leaf size, and color can be extracted from RGB images; the vegetation index and chlorophyll index can be obtained from hyperspectral data; dry and fresh weight and transcriptomics data can be obtained from the plant materials after this phenotyping. In addition, the WinRoots system can monitor crops in real time with sensors (such as RGB cameras and hyperspectral imaging spectrometers) and record phenotypic changes over time (**Figure 4B**) to analyze the dynamics of crop phenotypes. In conclusion, the WinRoots system allows the acquisition of multiple types of phenotypic data in a high-throughput manner from whole plants under controlled soil stress conditions and as such is well suited to meet the needs of high-throughput crop phenomics studies under soil stress.

Perspective and Challenges in Using the WinRoots System

The WinRoots system delivers high-throughput cultivation of crops under controlled soil stress conditions, as well as the instant acquisition of high-throughput and high-quality whole-plant phenotypic data. It will help integrate phenotypes, genotypes, and other omics data sources in a holistic approach to decipher the genetic regulatory networks and mechanisms underlying soil-stress-related traits and hasten precision crop breeding. Compared to previously reported high-throughput crop cultivation and phenotyping systems, the WinRoots system is flexible and offers controllable soil stress settings, while also saving space and cost. It has a wide potential applicability in medium-size crops such as wheat (*Triticum aestivum*), soybean, and rice, as well as large crops such as maize (*Zea mays*),

because it can be configured to different sizes and specifications of visual incubators.

The WinRoots system can obtain multiple phenotypic data by simulating soil stress conditions. In the future, it should be possible to integrate WinRoots with field phenotypes to obtain more production-guiding phenotypic indicators and thus speed up phenotypic screening and shorten the breeding cycle. Besides RGB cameras and hyperspectral imaging spectrometers, the WinRoots system can carry more sensors, such as chlorophyll fluorometers and infrared thermal imaging cameras, to expand the types of phenotypes being recorded. Additionally, the development of automatic image-based phenotyping analysis software is necessary to accommodate the volume of data generated, using machine learning methods to improve the efficiency of phenotypic analysis. The final version of this high-throughput and highly efficient phenotyping platform for crop phenomics studies under soil stress will greatly promote the development of plants with resistance to soil stress in crop breeding programs and related fundamental research.

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

YZ performed most of the experiments, data analysis and manuscript writing. WZ performed part of the experiments, data analysis and manuscript writing. QC assisted with the phenotypic data collecting. XZ, JY, TX, WS, XD, LW, JW and FZ performed part of the experiments. FX and XZ assisted with setting up the system. SL designed the system and the experiment, supervised the experiments, data analysis and writing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was financially supported by the National Natural Science Foundation of China (grant nos. U1906203 and 32072085), the National Key Research and Development Program of China (grant no. 2016YFD0101902), the National Key Research and Development Program of Shandong (grant no. 2021LZGC003), and the National Transgenic Project of China (grant nos. 2018ZX08009-14B and 2016ZX08010002-009).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2021.794020/full#supplementary-material>

REFERENCES

- Adu, M. O., Chatot, A., Wiesel, L., Bennett, M. J., Broadley, M. R., White, P. J., et al. (2014). A scanner system for high-resolution quantification of variation in root growth dynamics of *Brassica rapa* genotypes. *J. Exp. Bot.* 65, 2039–2048. doi: 10.1093/jxb/eru048
- Al-Tamimi, N., Brien, C., Oakey, H., Berger, B., Saade, S., Ho, Y. S., et al. (2016). Salinity tolerance loci revealed in rice using high-throughput non-invasive phenotyping. *Nat. Commun.* 17:13342. doi: 10.1038/ncomms13342
- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.tplants.2018.02.001
- Atkinson, J. A., Wingen, L. U., Griffiths, M., Pound, M. P., Gaju, O., Foulkes, M. J., et al. (2015). Phenotyping pipeline reveals major seedling root growth QTL in hexaploid wheat. *J. Exp. Bot.* 66, 2283–2292. doi: 10.1093/jxb/erv006
- Awlia, M., Nigro, A., Faikus, J., Schmoedel, S. M., Negrao, S., Santelia, D., et al. (2016). High-throughput non-destructive Phenotyping of traits that contribute to salinity tolerance in *Arabidopsis thaliana*. *Front. Plant Sci.* 7:1414. doi: 10.3389/fpls.2016.01414
- Bengough, A. G., Gordon, D. C., Al-Menaie, H., Ellis, R. P., Allan, D., Keith, R., et al. (2004). Gel observation chamber for rapid screening of root traits in cereal seedlings. *Plant Soil* 262, 63–70. doi: 10.1023/b:plso.0000037029.82618.27
- Bevan, M. W., Uauy, C., Wulff, B. B., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature* 543, 346–354. doi: 10.1038/nature22011
- Bolger, A. M., Poorter, H., Dumschott, K., Bolger, M. E., Arend, D., Osorio, S., et al. (2019). Computational aspects underlying genome to phenome analysis in plants. *Plant J.* 97, 182–198. doi: 10.1111/tpj.14179
- Busch, W., Moore, B. T., Martsberger, B., Mace, D. L., Twigg, R. W., Jung, J., et al. (2012). A microfluidic device and computational platform for high-throughput live imaging of gene expression. *Nat. Methods* 9, 1101–1106. doi: 10.1038/nmeth.2185
- Campbell, M. T., Knecht, A. C., Berger, B., Brien, C. J., Wang, D., and Walia, H. (2015). Integrating image-based phenomics and association analysis to dissect the genetic architecture of temporal salinity responses in rice. *Plant Physiol.* 168, 1476–89. doi: 10.1104/pp.15.00450
- Chen, Z. H., Zhou, M. X., Newman, I. A., Mendham, N. J., Zhang, G. P., and Shabala, S. (2007). Potassium and sodium relations in salinized barley tissues as a basis of differential salt tolerance. *Funct. Plant Biol.* 34, 150–162. doi: 10.1071/FP06237
- Choudhury, S. D., Samal, A., and Awada, T. (2019). Leveraging image analysis for high-throughput plant phenotyping. *Front. Plant Sci.* 10:508. doi: 10.3389/fpls.2019.00508
- Clark, R. T., Famoso, A. N., Zhao, K. Y., Shaff, J. E., Craft, E. J., Bustamante, C. D., et al. (2013). High-throughput two-dimensional root system phenotyping platform facilitates genetic analysis of root growth and development. *Plant Cell Environ.* 36, 454–466. doi: 10.1111/j.1365-3040.2012.02587.x
- Clark, R. T., MacCurdy, R. B., Jung, J. K., Shaff, J. E., McCouch, S. R., Aneshansley, D. J., et al. (2011). Three-dimensional root phenotyping with a novel imaging and software platform. *Plant Physiol.* 156, 455–465. doi: 10.1104/pp.110.169102
- Crowell, S., Korniliev, P., Falcao, A., Ismail, A., Gregorio, G., Mezey, J., et al. (2016). Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat. Commun.* 7:10527. doi: 10.1038/ncomms10527
- Cui, D. Z., Wu, D. D., Somarathna, Y., Xu, C. Y., Li, S., Li, P., et al. (2015). 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
- de Dorlodot, S., Forster, B., Pages, L., Price, A., Tuberosa, R., and Draye, X. (2007). Root system architecture: opportunities and constraints for genetic improvement of crops. *Trends Plant Sci.* 12, 474–481. doi: 10.1016/j.tplants.2007.08.012
- Edwards, D., and Batley, J. (2004). Plant bioinformatics: from genome to phenome. *Trends Biotechnol.* 22, 232–237. doi: 10.1016/j.tibtech.2004.03.002
- Furbank, R. T., and Tester, M. (2011). Phenomics - technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005
- Ghanem, M. E., Marrou, H., and Sinclair, T. R. (2015). Physiological phenotyping of plants for crop improvement. *Trends Plant Sci.* 20, 139–144. doi: 10.1016/j.tplants.2014.11.006
- Gioia, T., Galinski, A., Lenz, H., Muller, C., Lentz, J., Heinz, K., et al. (2017). GrowScreen-PaGe, a non-invasive, high-throughput phenotyping system based on germination paper to quantify crop phenotypic diversity and plasticity of root traits under varying nutrient supply. *Funct. Plant Biol.* 44, 76–93. doi: 10.1071/FP16128
- Gregory, P. J., Bengough, A. G., Grinev, D., Schmidt, S., Thomas, W. T. B., Wojciechowski, T., et al. (2009). Root phenomics of crops: opportunities and challenges. *Funct. Plant Biol.* 36, 922–929. doi: 10.1071/FP09150
- Grosskinsky, D. K., Svendsgaard, J., Christensen, S., and Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *J. Exp. Bot.* 66, 5429–5440. doi: 10.1093/jxb/erv345
- Hargreaves, C. E., Gregory, P. J., and Bengough, A. G. (2009). Measuring root traits in barley (*Hordeum vulgare* ssp. vulgare and ssp. spontaneum) seedlings using gel chambers, soil sacs and X-ray microtomography. *Plant Soil* 316, 285–297.
- Ho, M. D., McCannan, B. C., and Lynch, J. P. (2004). Optimization modeling of plant root architecture for water and phosphorus acquisition. *J. Theor. Biol.* 226, 331–340. doi: 10.1016/j.jtbi.2003.09.011
- in 't Zandt, D., Le Marie, C., Kirchgessner, N., Visser, E. J. W., and Hund, A. (2015). High-resolution quantification of root dynamics in split-nutrient rhizosides reveals rapid and strong proliferation of maize roots in response to local high nitrogen. *J. Exp. Bot.* 66, 5507–5517. doi: 10.1093/jxb/erv307
- Ingram, P. A., Zhu, J. M., Shariff, A., Davis, I. W., Benfey, P. N., and Elich, T. (2012). High-throughput imaging and analysis of root system architecture in *Brachypodium distachyon* under differential nutrient availability. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 1559–1569. doi: 10.1098/rstb.2011.0241
- Jie, L. (2015). Metabolite-based genome-wide association studies in plants. *Curr. Opin. Plant Biol.* 24, 31–38. doi: 10.1016/j.pbi.2015.01.006
- Junker, A., Muraya, M. M., Weigelt-Fischer, K., Arana-Ceballos, F., Klukas, C., Melchinger, A. E., et al. (2014). Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front. Plant Sci.* 5:770. doi: 10.3389/fpls.2014.00770
- Khan, M. A., Asaf, S., Khan, A. L., Jan, R., Kang, S. M., Kim, K. M., et al. (2019). Rhizobacteria AK1 remediates the toxic effects of salinity stress via regulation of endogenous phytohormones and gene expression in soybean. *Biochem. J.* 476, 2393–2409. doi: 10.1042/BCJ20190435
- Le Marie, C., Kirchgessner, N., Marschall, D., Walter, A., and Hund, A. (2014). Rhizosides: paper-based growth system for non-destructive, high throughput phenotyping of root development by means of image analysis. *Plant Methods* 10:13. doi: 10.1186/1746-4811-10-13
- Li, B., Chen, L., Sun, W., Wu, D., Wang, M., Yu, Y., et al. (2020). Phenomics-based GWAS analysis reveals the genetic architecture for drought resistance in cotton. *Plant Biotechnol. J.* 18, 2533–2544. doi: 10.1111/pbi.13431
- Li, M., Shao, M. R., Zeng, D., Ju, T., Kellogg, E. A., and Topp, C. N. (2020). Comprehensive 3D Phenotyping reveals continuous morphological variation across genetically diverse sorghum inflorescences. *New Phytol.* 226, 1873–1885. doi: 10.1111/nph.16533
- Li, Y. F., Zheng, Y., Vemireddy, L. R., Panda, S. K., Jose, S., Ranjan, A., et al. (2018). Comparative transcriptome and translome analysis in contrasting rice genotypes reveals differential mRNA translation in salt-tolerant Pokkali under salt stress. *BMC Genomics* 19:935. doi: 10.1186/s12864-018-5279-4
- Liu, H. J., Wang, F., Xiao, Y. J., Tian, Z. L., Wen, W. W., Zhang, X. H., et al. (2016). MODEM: multi-omics data envelopment and mining in maize. *Database (Oxford)* 2016:baw117. doi: 10.1093/database/baw117
- Luo, M. J., Zhao, Y. X., Zhang, R. Y., Xing, J. F., Duan, M. X., Li, J. N., et al. (2017). Mapping of a major QTL for salt tolerance of mature field-grown maize plants based on SNP markers. *BMC Plant Biol.* 17:140. doi: 10.1186/s12870-017-1090-7
- Mairhofer, S., Zappala, S., Tracy, S., Sturrock, C., Bennett, M. J., Mooney, S. J., et al. (2013). Recovering complete plant root system architectures from soil via X-ray μ -Computed Tomography. *Plant Methods* 9:8. doi: 10.1186/1746-4811-9-8
- Mano, Y., and Takeda, K. (1997). Mapping quantitative trait loci for salt tolerance at germination and the seedling stage in barley (*Hordeum vulgare* L.). *Euphytica* 94, 263–272.
- Manschadi, A. M., Christopher, J., Devoil, P., and Hammer, G. L. (2006). The role of root architectural traits in adaptation of wheat to water-limited environments. *Funct. Plant Biol.* 33, 823–837. doi: 10.1071/FP06055
- Manschadi, A. M., Hammer, G. L., Christopher, J. T., and deVoil, P. (2008). Genotypic variation in seedling root architectural traits and implications for

- drought adaptation in wheat (*Triticum aestivum* L.). *Plant Soil* 303, 115–129. doi: 10.1007/s11104-007-9492-1
- Mcdonald, G. K., Taylor, J. D., Verbyla, A., and Kuchel, H. (2017). Assessing the importance of subsoil constraints to yield of wheat and its implications for yield improvement. *Crop Pasture Sci.* 63, 1043–1065.
- Meister, R., Rajani, M. S., Ruzicka, D., and Schachtman, D. P. (2014). Challenges of modifying root traits in crops for agriculture. *Trends Plant Sci.* 19, 779–788. doi: 10.1016/j.tplants.2014.08.005
- Metzner, R., Eggert, A., van Dusschoten, D., Pflugfelder, D., Gerth, S., Schurr, U., et al. (2015). Direct comparison of MRI and X-ray CT technologies for 3D imaging of root systems in soil: potential and challenges for root trait quantification. *Plant Methods* 11:17. doi: 10.1186/s13007-015-0060-z
- Miransari, M. (2010). Contribution of arbuscular mycorrhizal symbiosis to plant growth under different types of soil stress. *Plant Biol. (Stuttg)* 12, 563–569. doi: 10.1111/j.1438-8677.2009.00308.x
- Morton, M. J. L., Awlia, M., Al-Tamimi, N., Saade, S., Pailles, Y., Negrao, S., et al. (2019). Salt stress under the scalpel - dissecting the genetics of salt tolerance. *Plant J.* 97, 148–163. doi: 10.1111/tpj.14189
- Muthamilarasan, M., Singh, N. K., and Prasad, M. (2019). Multi-omics approaches for strategic improvement of stress tolerance in underutilized crop species: a climate change perspective. *Adv. Genet.* 103, 1–38. doi: 10.1016/bs.adgen.2019.01.001
- Nagel, K. A., Lenz, H., Kastenholz, B., Gilmer, F., Aversch, A., Putz, A., et al. (2020). The platform GrowScreen-Agar enables identification of phenotypic diversity in root and shoot growth traits of agar grown plants. *Plant Methods* 16:89. doi: 10.1186/s13007-020-00631-3
- Nagel, K. A., Putz, A., Gilmer, F., Heinz, K., Fischbach, A., Pfeifer, J., et al. (2012). GROWSCREEN-Rhizo is a novel phenotyping robot enabling simultaneous measurements of root and shoot growth for plants grown in soil-filled rhizotrons. *Funct. Plant Biol.* 39, 891–904. doi: 10.1071/FP12023
- Perret, J. S., Al-Belushi, M. E., and Deadman, M. (2007). Non-destructive visualization and quantification of roots using computed tomography. *Soil Biol. Biochem.* 39, 391–399.
- Pfeifer, J., Kirchgessner, N., Colombi, T., and Walter, A. (2015). Rapid phenotyping of crop root systems in undisturbed field soils using X-ray computed tomography. *Plant Methods* 28:41. doi: 10.1186/s13007-015-0084-4
- Reynolds, M., Dreccer, F., and Trethowan, R. (2007). Drought-adaptive traits derived from wheat wild relatives and landraces. *J. Exp. Bot.* 58, 177–186. doi: 10.1093/jxb/erl250
- Rogers, E. D., Monastkova, D., Mijar, M., Nori, A., Goldman, D. I., and Benfey, P. N. (2016). X-Ray computed tomography reveals the response of root system architecture to soil texture. *Plant Physiol.* 171, 2028–2040. doi: 10.1104/pp.16.00397
- Siao, W., Coskun, D., Baluka, F., Kronzucker, H. J., and Xu, W. (2020). Root-apex proton fluxes at the centre of soil-stress acclimation. *Trends Plant Sci.* 25, 794–804. doi: 10.1016/j.tplants.2020.03.002
- Tardieu, F., Cabrera-Bosquet, L., Pridmore, T., and Bennett, M. (2017). Plant phenomics, from sensors to knowledge. *Curr. Biol.* 27, R770–R783. doi: 10.1016/j.cub.2017.05.055
- Tisne, S., Serrand, Y., Bach, L., Gilbault, E., Ben Ameer, R., Balasse, H., et al. (2013). Phenoscope: an automated large-scale phenotyping platform offering high spatial homogeneity. *Plant J.* 74, 534–544. doi: 10.1111/tpj.12131
- Tracy, S. R., Roberts, J. A., Black, C. R., McNeill, A., Davidson, R., and Mooney, S. J. (2010). The X-factor: visualizing undisturbed root architecture in soils using X-ray computed tomography. *J. Exp. Bot.* 61, 311–313. doi: 10.1093/jxb/erp386
- van Dusschoten, D., Metzner, R., Kochs, J., Postma, J. A., Pflugfelder, D., Buhler, J., et al. (2016). Quantitative 3D analysis of plant roots growing in soil using magnetic resonance imaging. *Plant Physiol.* 170, 1176–1188. doi: 10.1104/pp.15.01388
- Wu, H. H., Shabala, L., Zhou, M. X., Stefano, G., Pandolfi, C., Mancuso, S., et al. (2015). Developing and validating a high-throughput assay for salinity tissue tolerance in wheat and barley. *Planta* 242, 847–857. doi: 10.1007/s00425-015-2317-1
- Wu, J., Wu, Q., Pages, L., Yuan, Y. Q., Zhang, X. L., Du, M. W., et al. (2018). RhizoChamber-Monitor: a robotic platform and software enabling characterization of root growth. *Plant Methods* 14:44. doi: 10.1186/s13007-018-0316-5
- Wu, S., Wen, W., Wang, Y., Fan, J., and Guo, X. (2020). MVS-Pheno: a portable and low-cost phenotyping platform for maize shoots using multiview stereo 3D reconstruction. *Plant Phenomics* 2020:1848437. doi: 10.34133/2020/1848437
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087
- Yang, Y., and Guo, Y. (2018). Unraveling salt stress signaling in plants. *J. Integr. Plant Biol.* 60, 796–804.
- York, L. M., Nord, E. A., and Lynch, J. P. (2013). Integration of root phenes for soil resource acquisition. *Front. Plant Sci.* 4:355. doi: 10.3389/fpls.2013.00355
- Zhu, J., Ingram, P. A., Benfey, P. N., and Elich, T. (2011). From lab to field, new approaches to phenotyping root system architecture. *Curr. Opin. Plant Biol.* 14, 310–317. doi: 10.1016/j.pbi.2011.03.020
- Zhu, J., Kaeppler, S. M., and Lynch, J. P. (2005). Topsoil foraging and phosphorus acquisition efficiency in maize (*Zea mays*). *Funct. Plant Biol.* 32, 749–762. doi: 10.1071/FP05005
- Zorb, C., Geilfus, C. M., and Dietz, K. J. (2019). Salinity and crop yield. *Plant Biol. (Stuttg)* 21, 31–38.

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, Zhang, Cao, Zheng, Yang, Xue, Sun, Du, Wang, Wang, Zhao, Xiang 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.



Deciphering Novel Transcriptional Regulators of Soybean Hypocotyl Elongation Based on Gene Co-expression Network Analysis

Zhikang Shen¹ and Min Chen^{2,3*}

¹ College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China, ² State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, China, ³ Academy for Advanced Interdisciplinary Studies, Henan University, Kaifeng, China

OPEN ACCESS

Edited by:

Xianzhong Feng,
Northeast Institute of Geography and
Agroecology (CAS), China

Reviewed by:

Dior Rose Kelley,
Iowa State University, United States
Feng Ning Xiang,
Shandong University, China

*Correspondence:

Min Chen
chenmin@henu.edu.cn

Specialty section:

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

Received: 16 December 2021

Accepted: 17 January 2022

Published: 22 February 2022

Citation:

Shen Z and Chen M (2022)
Deciphering Novel Transcriptional
Regulators of Soybean Hypocotyl
Elongation Based on Gene
Co-expression Network Analysis.
Front. Plant Sci. 13:837130.
doi: 10.3389/fpls.2022.837130

Hypocotyl elongation is the key step of soybean seed germination, as well an important symbol of seedling vitality, but the regulatory mechanisms remain largely elusive. To address the problem, bioinformatics approaches along with the weighted gene co-expression network analysis (WGCNA) were carried out to elucidate the regulatory networks and identify key regulators underlying soybean hypocotyl elongation at transcriptional level. Combining results from WGCNA, yeast one hybridization, and phenotypic analysis of transgenic plants, a cyan module significantly associated with hypocotyl elongation was discerned, from which two novel regulatory submodules were identified as key candidates underpinning soybean hypocotyl elongation by modulating auxin and light responsive signaling pathways. Taken together, our results constructed the regulatory network and identified novel transcriptional regulators of soybean hypocotyl elongation based on WGCNA, which provide new insights into the global regulatory basis of soybean hypocotyl elongation and offer potential targets for soybean improvement to acquire cultivars with well-tuned hypocotyl elongation and seed germination vigor.

Keywords: gene co-expression network, hypocotyl elongation, auxin, light, transcription factor, soybean

INTRODUCTION

The consistency of seed germination and emergence is vital for large-scale agriculture, which not only reduces production cost but also increase yield potential. For dicots like soybean, hypocotyl elongation, the later stage of seed germination, determines whether the seedlings can break the soil smoothly and simultaneously, thus affecting the field emergence rate and consistency. Lots of research effort has focused on deciphering the regulation mechanism of hypocotyl development, mainly in the model species *Arabidopsis thaliana*. Similar to soybean, hypocotyl elongation in *Arabidopsis thaliana* is mainly driven by cell elongation. Effects of external environmental factors (light, temperature, gravity, etc.) as well endogenous hormones on hypocotyl elongation were well-studied (Deepika and Singh, 2020; Jiang et al., 2020). Generally speaking, the current research results show that the internal and external signals are integrated through PHYTOCHROME INTERACTING FACTORS (PIFs) to

regulate hypocotyl elongation (Leivar et al., 2008; Shen et al., 2008; Castillon et al., 2009; Liu et al., 2011; Wang and Wang, 2015; Kim et al., 2016).

Before the seeds germinate and break the soil, they experience the process of hypocotyl elongation, curved top and leaf yellowing (chloroplast undeveloped), which is called “skotomorphogenesis.” Among them, hypocotyl elongation plays a key role in seedling emergence (Alabadi et al., 2004; Josse and Halliday, 2008; Yi et al., 2020). Once exposed to light, hypocotyl elongation is inhibited. So far, a variety of photoreceptors have been found to be involved in the regulation of hypocotyl elongation (Yu et al., 2013; Srivastava et al., 2015; Hu et al., 2021; Lin et al., 2021; Zhang et al., 2021; Zhong et al., 2021). For example, mutations of phytochrome A (PHYA) and phytochrome B (PHYB) caused insensitivity to far-red and red light in terms of the inhibition of hypocotyl elongation response (Seo et al., 2009; Lee et al., 2012; Zhong et al., 2012), indicating that PHYA and PHYB mediate the inhibition of hypocotyl elongation by far-red light and red light, respectively. Similarly, cryptochrome *cry1* mutant showed resistance to blue light induced inhibition of hypocotyl elongation, indicating that CRY1 mediated the inhibitory effect of blue light on hypocotyl elongation.

PIFs are basic helix-loop-helix (bHLH) transcription factors (TFs) that can interact with phytochrome proteins. They contain APA motifs which are capable of binding to PHYA or APB motifs binding to PHYB. So far, it has been reported that at least seven PIFs (1, 3, 4, 5, 6, 7, and 8) can interact with PHYA or PHYB (Leivar et al., 2009, 2012; Shin et al., 2009; Nozue et al., 2011; Hornitschek et al., 2012; Li et al., 2012; Oh et al., 2012; Jeong and Choi, 2013; Zhang et al., 2013). PHYA and PHYB can induce the phosphorylation and rapid degradation of their target PIFs through the interaction, while PIFs as TFs usually bind to the G-Box motif (CACGTG) in the promoter region of their downstream targets, such as *HISTONE DEACETYLASE 15 (HDA15)* and *TIMING OF CAB EXPRESSION1 (TOC1)*, and regulate their expression to promote hypocotyl elongation. Therefore, light signal promotes the degradation of PIFs by activating the activity of PHYA/PHYB, so as to inhibit hypocotyl elongation. Interestingly, PIFs are also involved in multiple hormone signaling pathways to control hypocotyl elongation. For example, PIF3 is activated by ETHYLENE INSENSITIVE3 (EIN3) and EIN3-LIKE1 (EIL1) in ethylene signaling pathway, as well act downstream of gibberellin (GA) signaling pathway, to promote hypocotyl elongation. Therefore, PIFs are the key regulators of hypocotyl elongation (Hyun and Lee, 2006; Mara et al., 2010; Castelain et al., 2012; Sassi et al., 2013; Gommers et al., 2017).

Almost all plant hormones are directly or indirectly involved in the regulation of hypocotyl elongation (Reed et al., 2018). It has been found that in *Arabidopsis thaliana*, auxin, GA or brassinosteroids (BRs) can promote the hypocotyl elongation, while abscisic acid (ABA) and cytokinin (CK) inhibit the process. For auxin, the expression of cell wall-related *EXPANSIN* genes was induced in hypocotyl grown under dark condition by auxin treatment (Pelletier et al., 2010; Cosgrove, 2016; Majda and Robert, 2018). A recent study showed that auxin induces

the interaction between transmembrane kinase (TMK) and H^+ -ATPases in the plasma membrane, which facilitate proton transport from intracellular to cell wall, where EXPANSINs are activated to loosen and expand cell wall structure to allow the cell elongating in hypocotyl (Hocq et al., 2017; Ivakov et al., 2017; Duman et al., 2020; Xin et al., 2020; Lin et al., 2021). The role of ethylene (ET) is slightly complicated: it inhibits hypocotyl elongation under dark conditions and promotes elongation under light conditions (Jin et al., 2021). Studies have shown that ethylene acts through EIN3 to activate two opposite pathways (Paque et al., 2014), one mediated by PIF3 (phytochrome interacting factor 3) to promote hypocotyl elongation under light, and the other through ERF1 (ETHYLENE RESPONSE FACTOR1) to inhibit hypocotyl elongation under dark.

Soybean (*Glycine max*) is one of the most important economic crops in the world because of its high oil and protein concentrations (Liu et al., 2020). Similar to *Arabidopsis*, the elongation of soybean hypocotyl is affected by environmental factors (temperature, light, soil hardness, etc.) and phytohormones (auxin, GA, BRs, etc.) (Wang et al., 2021), but the underlying molecular mechanisms remain largely unclear. Studies have shown that soybean and *Arabidopsis thaliana* may have similar mechanisms of hypocotyl elongation. For example, a recent study showed that low light and high temperature promote soybean hypocotyl elongation through elevating the biosynthesis of auxin and GA, which is similar to that in *Arabidopsis thaliana* (de Lucas et al., 2008; Bawa et al., 2020). However, soybean has a genome size of 1.1–1.15 GB, which is much larger than that of *Arabidopsis thaliana*, implying soybean has more complicated regulatory networks underlying hypocotyl elongation. In order to elucidate the global architecture of the regulatory networks of soybean hypocotyl elongation, this study adopted the weighted gene co-expression network analysis (WGCNA) integrated with bioinformatics approaches. WGCNA was designed to find clusters (modules) of highly related genes, and correlate modules with external sample traits, so as to identify candidate regulators. Our study integrated the results of various big data analysis and obtained two transcriptional regulatory submodules and multiple novel candidate genes for hypocotyl elongation regulation. Our findings provide suitable combination of genes for the use of molecular breeding to produce soybean varieties with high germination rate and yield potential in the field.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

For soybean, cultivar Williams 82 was used as research material unless specified. For *Arabidopsis thaliana*, Columbia-0 ecotype was used as WT. The *Arabidopsis* T-DNA insertion line were obtained from Arashare (<https://www.arashare.cn>). All *Arabidopsis* T-DNA insertion mutants used in this study were listed in **Supplementary Table 3**. The T-DNA insertion lines were verified according to the method provided by the Salk Institute (<http://signal.salk.edu/tdnaprimers.2.html>). The primers used for T-DNA insertion verification were listed in

Supplementary Table 6. All seeds were collected and stored under the same conditions.

For hypocotyl length analysis of *Arabidopsis* mutant lines, *Arabidopsis* seeds were surface-sterilized with 70% ethanol for 10 min, followed by four times washing with sterile-deionized water, and then grown on half-strength Murashige and Skoog (MS) medium (0.9% agar, pH 5.7), with or without a 16 h light/8 h dark photoperiod (light density $100 \mu\text{mol m}^{-2} \text{s}^{-1}$) at 22°C in a plant growth chamber. Hypocotyl length was determined in 5-day-old seedlings.

For phenotypic analysis of soybean seedlings, soybean seeds were surface-sterilized with 5% Sodium hypochlorite for 5 min, followed by four times washing with sterile-deionized water, and then germinated on filter papers for 2 d. The well-germinated seeds were planted in the vermiculite medium, and grown in a plant growth chamber with or without continuous white light (light density at $200 \mu\text{mol m}^{-2} \text{s}^{-1}$) at 27°C . Hypocotyl length was determined in 5-day-old seedlings.

Exogenous Application of Phytohormone

The uniform seedlings (3-day-old, grown under continuous white light) were selected for phytohormone treatment. Seedlings were treated with phytohormone as follows: $10 \mu\text{M}$ IAA (Sangon, Shanghai, China), or $100 \mu\text{M}$ Kyn (Merck, Darmstadt, Germany), or water as control. The hypocotyl samples were collected after 2 d treatment for hypocotyl length measurement (with Image-Pro Plus software) and 30 min/2 h for RNA extraction. All treatments were performed with at least three biological replicates.

RNA Extraction and RT-qPCR Analyses

Total RNA was extracted from at least 200 mg soybean hypocotyl using a plant RNAprep Kit (TianGen, Beijing, China). RNA concentration and purity were checked using the Qubit RNA Assay Kit using Qubit 2.0 Fluorometer (Life Technologies, CA, USA). First-strand complementary DNA (cDNA) was synthesized with $2 \mu\text{g}$ total RNA using a Reverse Transcription kit (Vazyme, Nanjing, China) and 80 ng of first strand cDNA were used as the template for qPCR. The qPCR analyses were performed on a CFX96 Touch Real-Time PCR Detected System (BIO-RAD, CA, USA). The RNA levels were calculated as described before (Chen and Penfield, 2018). The reference gene was a soybean ACTIN gene (*GLYMA_02G091900*). Three biological replicates with three technical replicates were performed for each sample. Primer sequences are listed in **Supplementary Table 6**.

RNA Sequencing

Total RNA was extracted from at least 250 mg soybean hypocotyl and root tip samples, respectively. RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). Sequencing libraries were generated with 400 ng total mRNA using the NEB RNA Library Prep Kit for Illumina (NEB, MA, USA), and for quality control using Agilent 2100 bioanalyzer. All libraries were sequenced in paired-end 150 bases protocol (PE150) on an Illumina NovaSeq 6000 system (Illumina, CA, USA).

Quality control of sequencing data was carried out with fastp. The raw reads were then trimmed to discard adapter sequences, reads containing poly-N and with low quality by using TrimGalore. After that, 6.19–9.12 GB clean reads were obtained for further analysis. Clean reads were mapped to the soybean reference genome (Williams 82 V2.1) using Hisat2 with default parameters. Number the reads mapped to each gene was calculated with featureCounts. And then fragments per kilobase of transcript per million mapped reads (FPKM) of each gene was calculated based on the length of the gene and reads count mapped to this gene using R. All raw reads were deposited to the National Center for Biotechnology Information Sequence Reads Archive (SRA) with accession number: PRJNA789395, PRJNA789276, and PRJNA790743.

Regulatory Network Construction by WGCNA

WGCNA was performed to identify gene co-expression networks associated with hypocotyl elongation by using the WGCNA software package as previously described (Langfelder and Horvath, 2008). Briefly, after excluding genes with missing entries or average FPKM < 1 , 18 samples and 23,729 genes were selected to construct the weighted gene co-expression network. The correlation matrix was then constructed using pairwise Pearson Correlations among all genes. The soft threshold was set to $\beta = 24$ to achieve a scale-free network. To identify gene modules, the dynamic tree-cutting algorithm was used. Gene modules with similar expression profiles were merged (PCC > 0.85). Gene connectivity in each module was determined, and hub genes were identified as the genes with the highest connectivity. The global gene co-expression network was visualized with Cytoscape_v3.5.1 (Shannon et al., 2003).

To identify modules correlated with hypocotyl elongation, a network-weighted gene significance function was used. Gene significance (GS) was defined as the correlation between a specific gene expression level and the hypocotyl development. Gene modules associated with hypocotyl elongation were also identified based on the correlation between Module eigengene (ME) and hypocotyl development. ME is considered representative expression of all genes in one module and was identified using the ME function.

To analyse functional annotation of the identified genes, all genes in the candidate modules were subjected to gene ontology (GO) functional annotation and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses using the ClusterProfiler tool and KEGG website (<https://www.genome.jp/kegg>) (Yu et al., 2012). Functional enrichment was performed in the following three GO categories: biological process (BP), molecular function (MF), and cellular component (CC).

Cis-Motif Enrichment Analysis

Motif discovery was performed with the MEME suite (<https://meme-suite.org/>), using the 2.5 kb sequence upstream of the putative target genes' ATG start codon. The TOMTOM tool was exerted to predict the functions of the enriched motifs by comparing them with known cis-motifs.

Creation of Transgenic Plants

To generate *GmPRE6* (*Glyma.01g044800* and *Glyma.08g298700*)-overexpression lines, the full-length coding sequence (CDS) of the two *GmPRE6s* was amplified and cloned into the pCambia1300-eCFP vector separately. The constructs were then introduced into *Agrobacterium* strain GV3101 and transformed into *Arabidopsis thaliana* plants by floral infiltration. The T1 transgenic plants were selected on hygromycin-containing medium, and the progeny were used for subsequent experiments.

Agroinfiltration of *N. benthamiana*

The pCambia1300-eCFP plasmid that harboring *GmPRE6* CDS (driven by the *CaMV* 35S promoter) were mobilized into GV3101. Recombinant *agrobacteria* were prepared for infiltration using a modified protocol published before. In short, recombinant *bacteria* cultures were incubated overnight at 28 °C with shaking. *Bacteria* were pelleted by centrifugation (14,000 g for 5 min) and resuspended to an OD600 = 1.0 in MMA (10 mM MES pH 5.6, 10 mM MgCl₂, 200 μM acetosyringone). *Bacteria* were delivered into the underside of young leaves using a blunt tipped plastic syringe and applying gentle pressure. *N. benthamiana* leaf samples were collected 24–48 h post agroinfiltration and the fluorescence was imaged using a Leica confocal microscope (LSM 900, Zeiss, Germany).

Yeast One-Hybrid Assay

Y1H assays were performed following a previously published method (Fuxman Bass et al., 2016) with minor modifications. Briefly, a 370-bp, a 200-bp, and a 643-bp promoter fragment of *Glyma.12G034600*, *Glyma.12G033900*, and *Glyma.11G031700* was amplified (Supplementary Table 6) and cloned into the pAbAi-bait vector, respectively. The pAbAi-bait plasmids were subsequently linearized and transformed into yeast (Y1H Gold strain). The transformed colonies were screened by colony PCR. The prey constructs (pGADT7-TFs) were transformed into yeast harboring a pAbAi-bait plasmid and screened on the SD-Leu/Aba plate. The minimum inhibitory concentration of aureobasidin A was 200 ng/mL. At least three independent biological replicates were performed for each combination.

RESULTS

A Hub Module Regulating Hypocotyl Elongation Identified Using WGCNA

In order to explore the global regulatory mechanisms underlying soybean hypocotyl elongation, we profiled the transcriptomes of the emerging hypocotyl and root tip tissue at 0, 1, and 5 days after imbibition (DAI) in the soybean cultivar “Williams 82” using RNA-seq (Supplementary Figure 1). The root tip tissue was set as negative control since its elongation mode is largely different from that of hypocotyl (root tip has vigorous cell division but hypocotyl does not). Eighteen samples, 23,729 genes were screened and analyzed using WGCNA. After normalization, no outlier module was eliminated. The soft-thresholding power was set to $\beta = 24$ (scale free $R^2 = 0.92$) to ensure a scale-free network (Supplementary Figure 2). We identified a total of 97 modules via average linkage hierarchical clustering (Figure 1A).

The number of genes contained in these clusters ranged from 58 to 9,854 (Supplementary Table 1), and some of them were highly correlated between each other (Figure 1B). In order to simplify the analysis, we combined clusters with Pearson correlation coefficient (PCC) >0.85, and finally obtained 19 modules for subsequent analysis (Figure 1C; Supplementary Table 1).

Genes in the same module could form networks and may participate in similar biological processes, so we examined the relationship between modules and hypocotyl elongation phenotype. By calculating the correlation significance between gene expression level in every single module and hypocotyl elongation phenotype, we found five modules with average significance higher than 0.5, of which a cyan module reached 0.94 (Figure 2A; Supplementary Table 2), indicating it may participate in hypocotyl elongation. Further, we analyzed the correlation between module eigengene (ME, which reflects the overall gene expression level of the entity module) and hypocotyl elongation, and again found that cyan module was significantly correlated with hypocotyl elongation phenotype ($R = 0.99$; Figure 2B). Therefore, the cyan module is considered to be the hub module regulating hypocotyl elongation. Consistent with the above results, most of the genes in the cyan module were specifically highly expressed in the hypocotyl tissue at 5 DAI (Supplementary Figure 3), and the connectivity of the genes was significantly positively correlated with its significance ($R = 0.91$, Figure 2C).

To confirm that cyan module is indeed a regulation module of soybean hypocotyl elongation, we randomly selected six genes with high connectivity in the module (*Glyma.14g218800*, *Glyma.15g253700*, *Glyma.03g060300*, *Glyma.15g209000*, *Glyma.06g039900*, *Glyma.16g205200*, and obtained their *Arabidopsis* homolog knockout mutants (homolog inferred from NCBI blast program, and when there are multiple results, the one with the highest score was selected; *Arabidopsis* mutants purchased from AraShare <https://www.arashare.cn>). We found that four out of six *Arabidopsis* mutants showed altered hypocotyl elongation compared to wild type control (*hxx3*, corresponding to *Glyma.14g218800*; *lhcb5*, corresponding to *Glyma.16g205200*; *lhcb4.2*, corresponding to *Glyma.03g060300*; *emb1473*, corresponding to *Glyma.15g209000*) (Supplementary Figure 4; Supplementary Table 3), indicating that the cyan module is a convincing regulatory candidate of soybean hypocotyl elongation.

Characterization of a Conserved PIFs-SAURs-EXPANSINs Submodule

There are 1,532 genes in the cyan module. Gene Ontology biological process (GO-BP) analysis showed that the genes in the cyan module were mainly enriched in GO-BP terms related to auxin response, light response and plant cell wall structure (Supplementary Figure 6). Among them, the genes related to cell wall structure contain 10 *EXPANSINs*, which might be involved in auxin induced cell wall expansion and hypocotyl elongation as previously reported (Zhao et al., 2008). These results were consistent with the previous reported mechanisms in *Arabidopsis thaliana*, which highlighted the regulation of

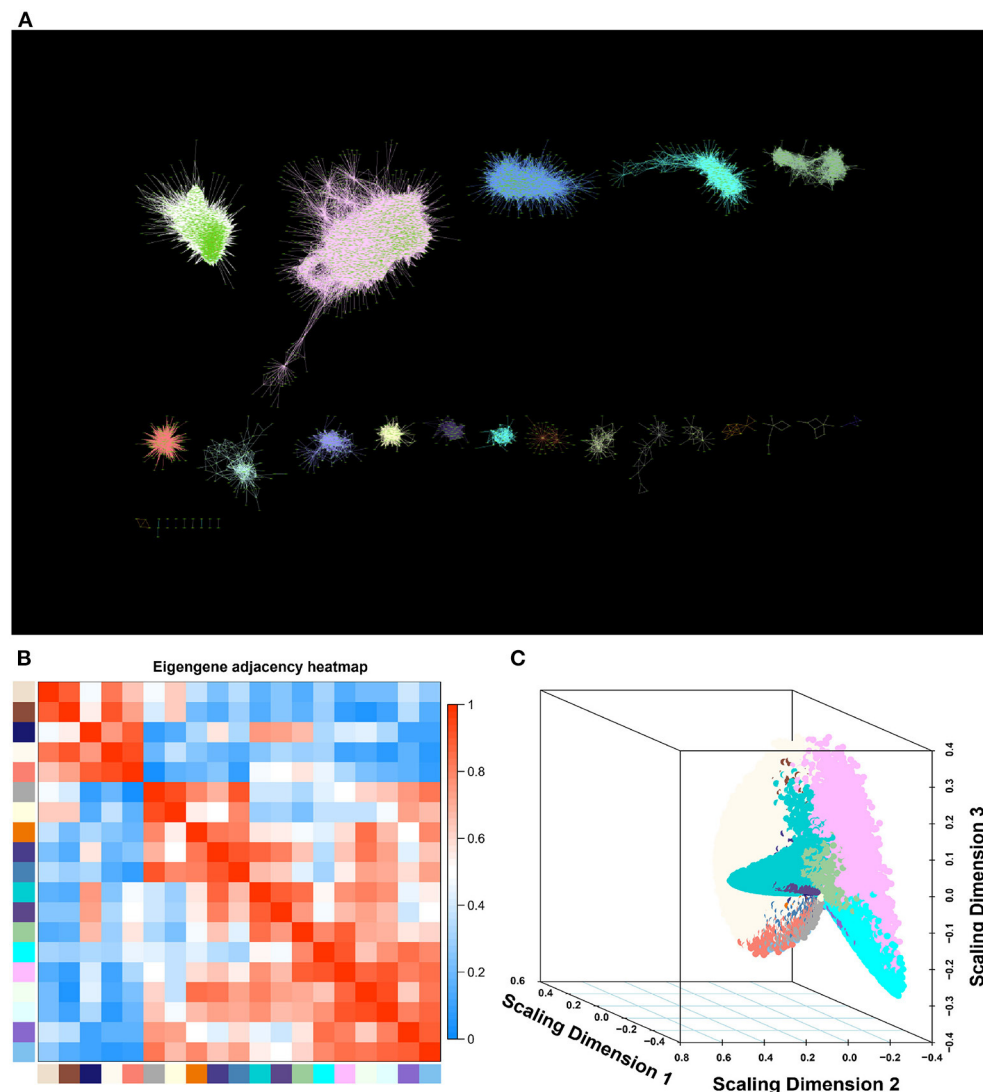


FIGURE 1 | A global gene co-expression network related to hypocotyl length in soybean. **(A)** Global gene co-expression network including 19 modules visualized with Cytoscape. **(B)** Eigengene adjacency heatmap of 19 modules. Each row/column corresponds to a module. Each cell contains a corresponding correlation indicated by color. **(C)** Multidimensional scaling plots of the TOM-based dissimilarity of gene expression in different modules.

hypocotyl elongation by auxin and light. Interestingly, when comparing the genes in cyan module with the four previously reported quantitative trait loci (QTLs) related to hypocotyl elongation (Lee et al., 2001; Liang et al., 2014), we found 27 genes appearing in both results (**Figures 3A,B**). To study whether and how these genes are involved in regulating hypocotyl elongation, the promoters of these genes and their homologous genes in *Arabidopsis thaliana* (homolog inferred from NCBI blast program; promoter spans 2-kb region upstream of the ATG translation start site) were scanned for cis-elements (MEME suite <https://meme-suite.org/>) (Bailey et al., 2009). Interestingly, the enriched elements included the one responsive to auxin (class I), the one responsive to light response (class II), etc. (**Supplementary Figure 7**), suggesting that these genes might be

involved in the auxin and light regulating hypocotyl elongation process. It is worth noting that in these genes there are five small auxin up RNAs (SAURs) members. Previous studies have shown that SAURs act as PIFs' direct targets to participate in hypocotyl elongation in *Arabidopsis thaliana*. In the cyan module, there are two members of soybean PIF family, *GmPIF1* (*Glyma.13g130100*) and *GmPIF3* (*Glyma.19g222000*) (weighted value = 0.48). We found that *GmPIF1* and *GmPIF3* were of high connectivity with these five *GmSAURs* as well one other auxin responsive gene *GmRGF6* (*ROOT MERISTEM GROWTH FACTOR6*, *Glyma.11G031700*) in the 27 genes (weighted value = 0.51). Intriguingly, the 10 *EXPANSIN* genes related to cell wall growth in the cyan module were closely correlated with *GmSAURs* and *GmRGF6*. Based on these results, we constructed

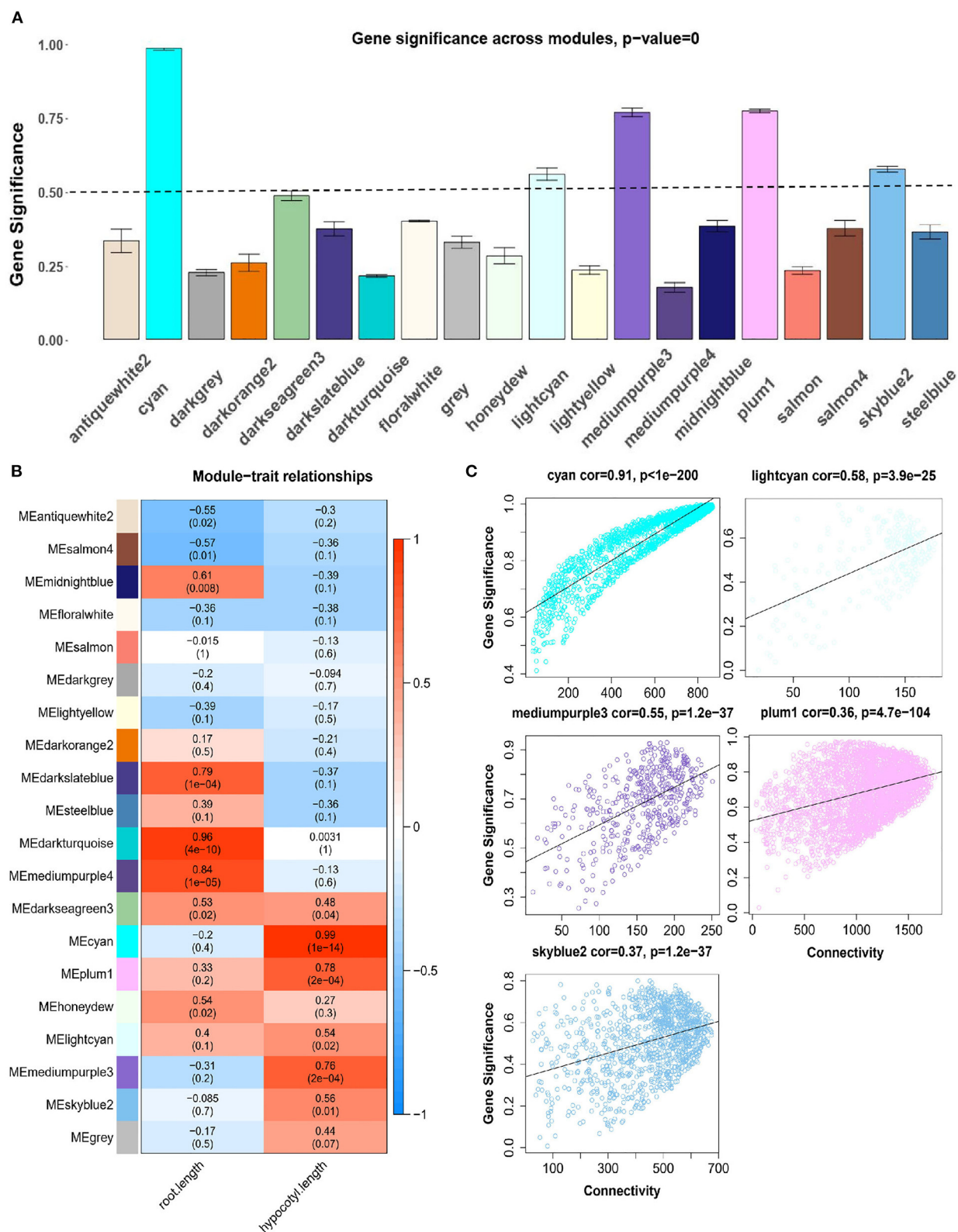
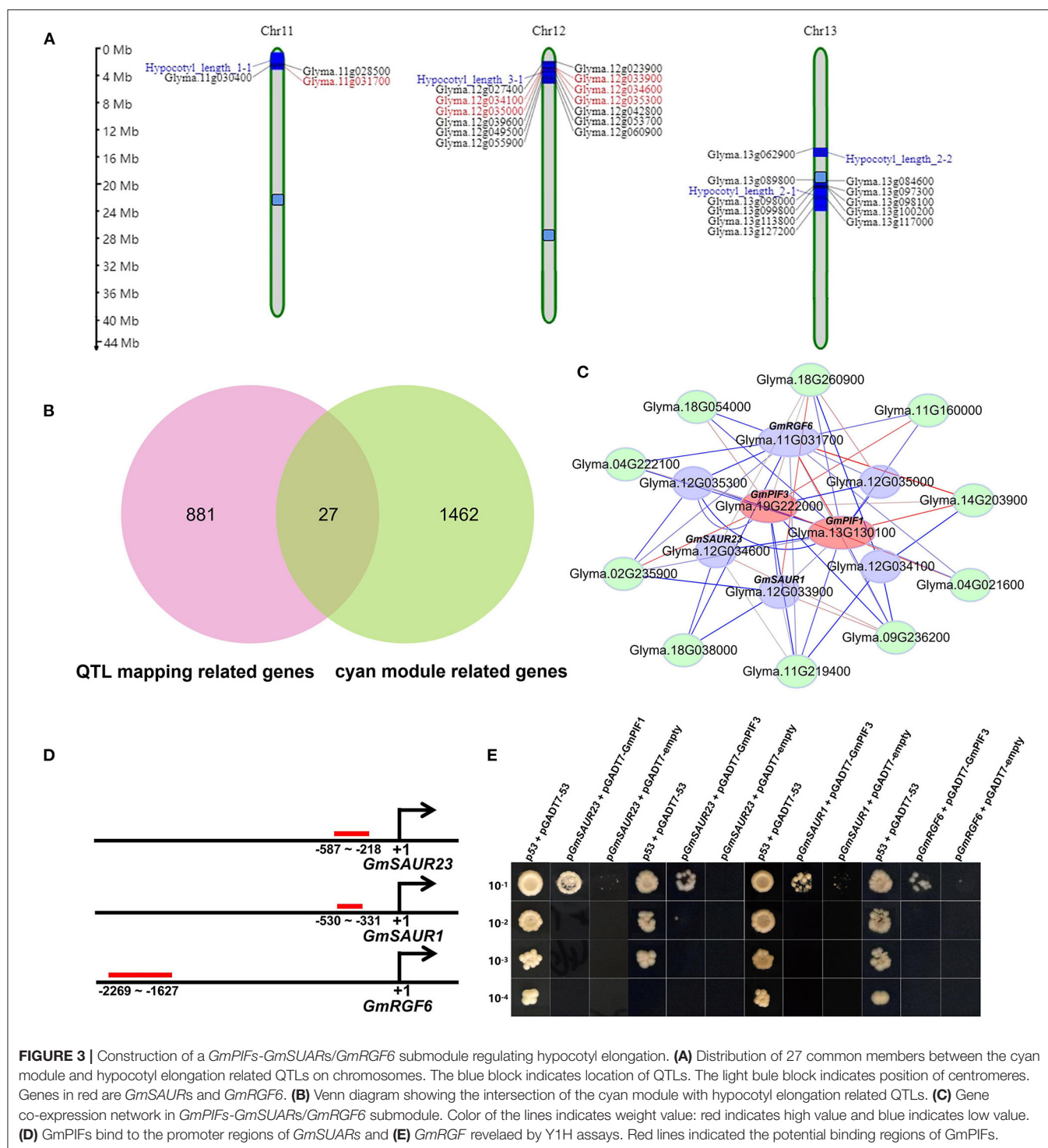


FIGURE 2 | Correlation between modules and traits. **(A)** The mean gene significance between module and hypocotyl development. **(B)** Module-trait relationship. Each row corresponds to a module, while each column corresponds to the root/hypocotyl development trait. Each cell contains a corresponding correlation. **(C)** The correlation between gene significance and connectivity in the 5 modules with gene significance > 0.5 in **(A)**.



a novel submodule composed of *GmPIF1/3*, *GmSAURs*, *GmRGF6* and *EXPANSINs* as the regulator of hypocotyl elongation (Figure 3C).

In *Arabidopsis thaliana* and other species, PIFs act through SAURs to regulate hypocotyl elongation in response to external and internal cues such as light, auxin and temperature. Through

the analysis of previously published ChIP-seq data (Pfeiffer et al., 2014), we found that both AtPIF1 and AtPIF3 bind to the promoter regions of AtSAUR1, AtSAUR23 and AtRGF6 (Supplementary Table 4). *GmPIF1* and *GmPIF3* have high homology to AtPIF1 and AtPIF3 (Supplementary Figure 5), suggesting that *GmPIF1* and *GmPIF3* may act through the

similar mechanism, i.e., binding to the promoters of *GmSAURs*. To test this hypothesis, the regions containing class I and II motifs (**Supplementary Figure 7**) in the promoter regions of *GmSAUR1*, *GmSAUR23*, and *GmRGF6* (**Figure 3D**) were selected as baits in yeast one hybridization analysis to identify interactors. Not surprisingly, it is found that *GmPIF1* can bind to the promoter of *GmSAUR23*, and *GmPIF3* can bind to the promoter of *GmSAUR23*, *GmSAUR1*, and *GmRGF6* (**Figure 3E**), indicating that the binding of PIFs to the promoter region of SAURs is indeed a conservative mechanism.

Auxin and light are important factors regulating hypocotyl elongation. Similar to the case for *Arabidopsis*, light inhibited the elongation of soybean hypocotyl (data not shown), while exogenous auxin promoted the elongation (**Supplementary Figure 8**). The above studies showed that PIFs-SAURs-EXPANSINs submodule may be involved in hypocotyl elongation regulated by light and auxin. To verify this, we examined the expression levels of *GmPIF1*, *GmPIF3*, *GmSAUR1*, *GmSAUR23*, and *GmRGF6* and two *EXPANSINs* (*Glyma.14g203900* and *Glyma.02g235900*) under dark-light change and exogenous auxin by RT-qPCR (**Figure 4**; **Supplementary Figure 9**). The results show that the expression of *GmPIF1* was induced by light, while the expression of *GmPIF3* was first induced and then inhibited by light. Previous studies have shown that light mainly regulates *AtPIF1* and *AtPIF3* at the protein level (Qiu et al., 2017). Our results suggest that light may regulate the expression of PIFs at the transcriptional level in soybean, which reflects a feedback regulation mechanism.

The expression of *GmSAUR1*, *GmSAUR23*, and *GmRGF6* was inhibited by light, while previous studies in *Arabidopsis thaliana* showed that light induced the expression of several other SAURs (Dong et al., 2019), suggesting that SAUR family members may have functional differentiation among different species. When treated with exogenous indole-3-acetic acid (IAA) and L-Kynurenine (Kyn), an auxin biosynthesis inhibitor, the expression of *GmPIF1* and *GmPIF3* in the hypocotyl showed slight responses, which may be due to auxin mainly inhibiting the transcriptional activation activity of PIFs through downstream AUX/IAAs rather than regulating their expression directly. The expression of *GmRGF6* was induced by exogenous IAA and inhibited by Kyn, which is similar to the case in *Arabidopsis thaliana* (Busatto et al., 2017). The expression of *GmSAUR1* and *GmSAUR23* was similarly induced by exogenous IAA, but with no obvious response to Kyn. It should be noted that the expression of the two *EXPANSIN* genes was inhibited by light, and induced by exogenous IAA (**Supplementary Figure 9**). These results suggest that *GmSAUR1*, *GmSAUR23*, and *GmRGF6* may be positive regulators of soybean hypocotyl elongation. They perceive signals such as light and auxin and regulate hypocotyl elongation through downstream EXPANSINs.

Characterization of a *GmPRE6s* Based Submodule

Genes with similar expression pattern in a single module might be targets of a cluster of TFs involved in the same biological process (Zaidi et al., 2020). In order to identify the

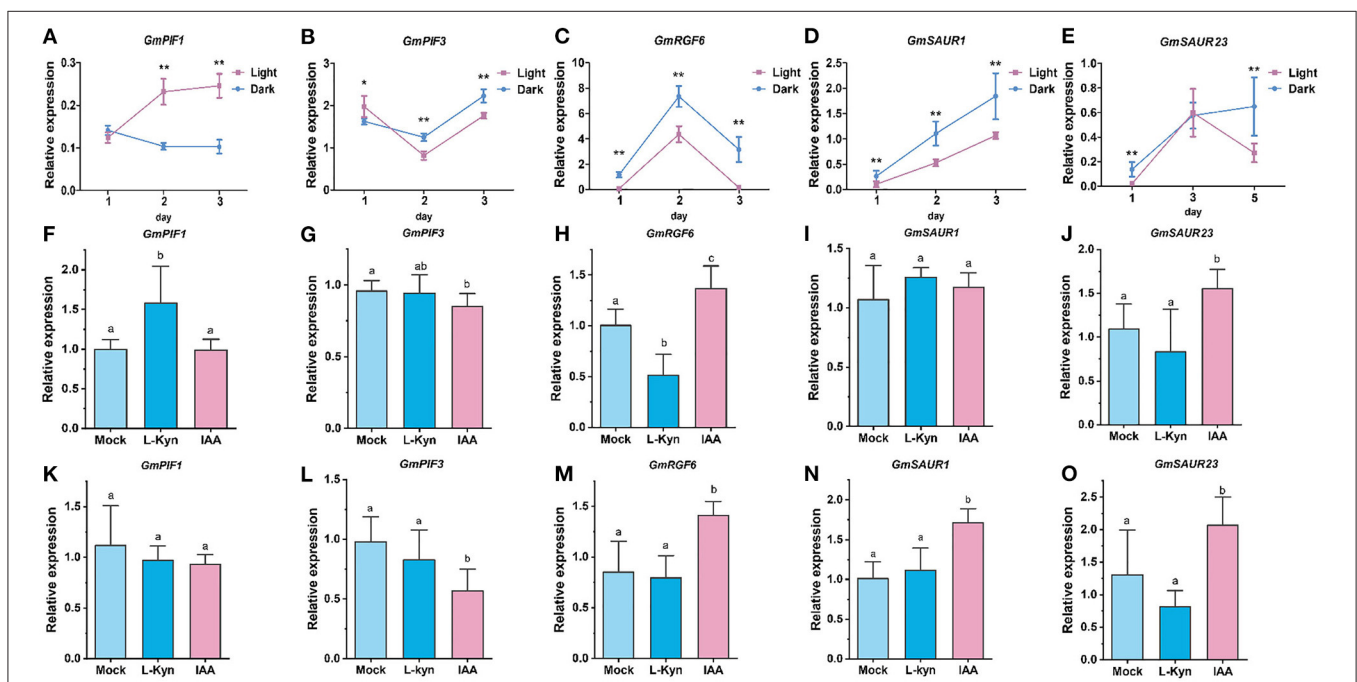
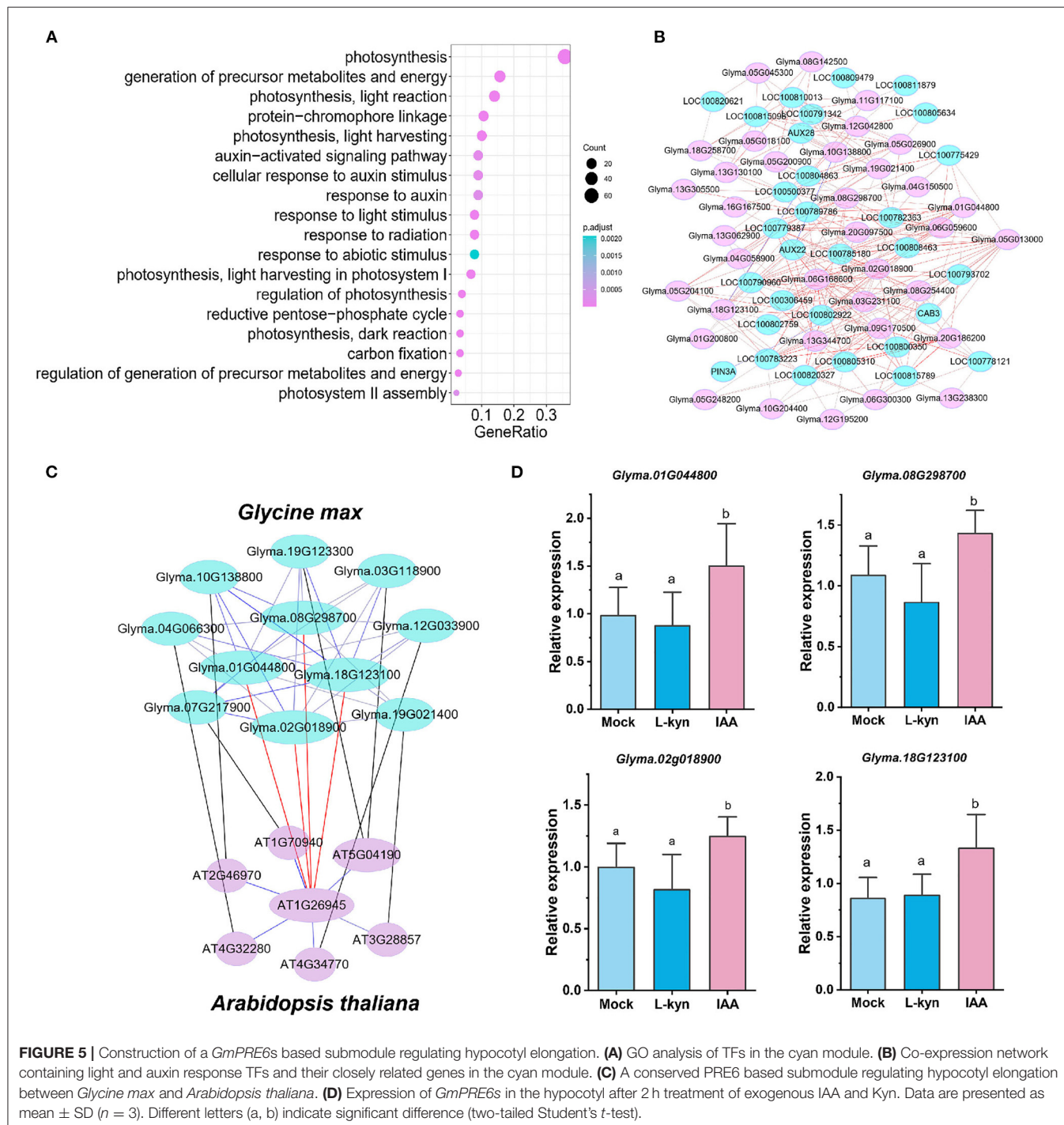


FIGURE 4 | Expression of *GmPIFs* and *GmSAURs/GmRGF6* in response to light and exogenous auxin. (A–E) *GmPIFs* and *GmSAURs/GmRGF6* expression in soybean hypocotyl under light and dark condition. (F–J) *GmPIFs* and *GmSAURs/GmRGF6* expression in the hypocotyl after 30 min treatment of exogenous IAA and Kyn. (K–O) *GmPIFs* and *GmSAURs/GmRGF6* expression in the hypocotyl after 2 h treatment of exogenous IAA and Kyn. Data are presented as mean \pm SD ($n = 3$). A statistically significant difference of $p < 0.05$ or $p < 0.01$ relative to mock (two-tailed Student's *t*-test) is denoted by * and **, respectively. Different letters (a, b) indicate significant difference (two-tailed Student's *t*-test).

TFs which potentially regulate the hypocotyl elongation in the cyan module, we used PlantTFDB (planttfdb.gao-lab.org) to predict the TFs, and obtained 83 TFs of 24 families in the cyan module. Among them, the bHLH family transcription factors (Supplementary Figure 10) accounted for the largest proportion (22.9%). Gene Ontology biological process (GO-BP) analysis showed that many of these TFs were involved in the response to light and auxin (36/83, Figure 5A). We extracted these transcription factors and 85 closely connected

genes to them (weighted value > 0.45) in the cyan module, and constructed a TF based submodule (Figure 5B). It is not surprising that the submodule includes multiple auxin responsive and light responsive genes (Supplementary Table 7), such as *PIN3A* (*AUXIN EFFLUX CARRIER COMPONENT3A*, *Glyma.07G217900*, involved in auxin polar transport), *AUX22* and *AUX28* (*AUXIN-INDUCED PROTEIN22* and *28*, *Glyma.08G207900*, and *Glyma.19G161100*, involved in auxin signal transduction) and *CAB3* (*CHLOROPHYLL A/B*



BINDING PROTEIN3, *Glyma.05G128000*, responsive to light) etc., indicating this submodule act as well in a light and auxin response dependent manner.

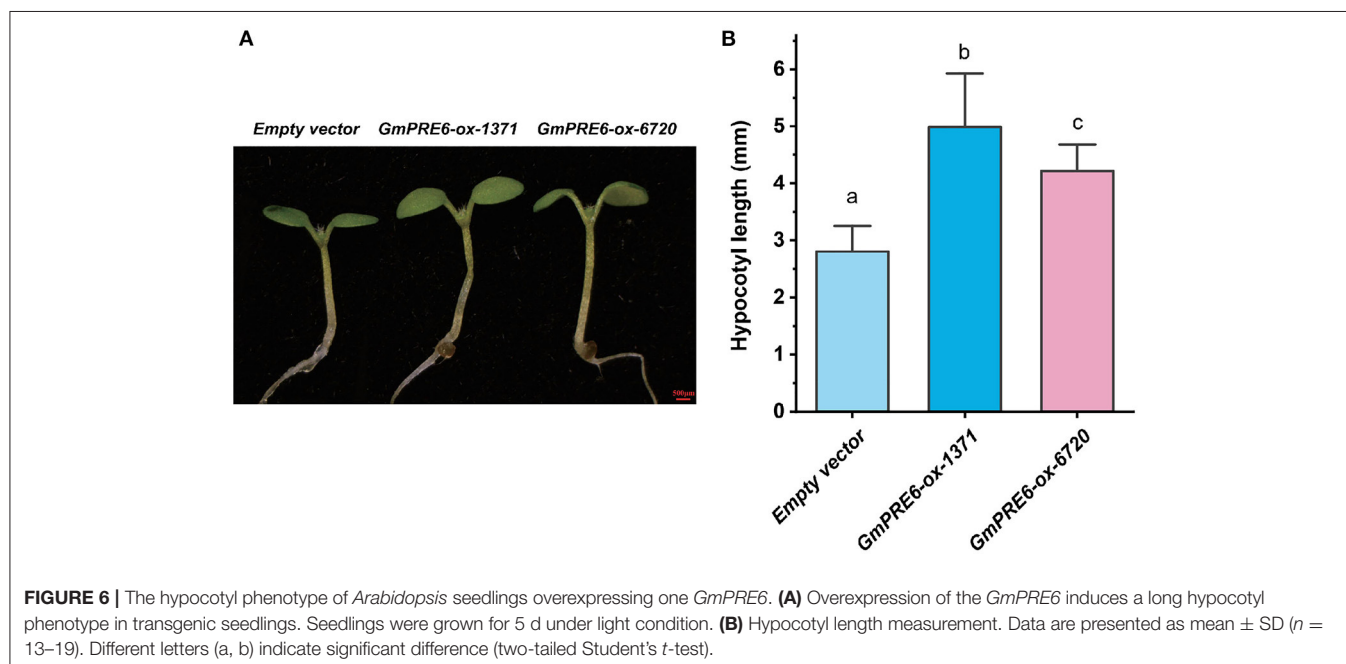
Interestingly, in this submodule, there are four homologous genes of *Arabidopsis* *PACLOBUTRAZOL RESISTANCE6* (*PRE6*). PREs are atypical bHLH transcription factors that lack the basic DNA binding domain. Accumulated evidence suggests that PREs are involved in the regulation of hypocotyl elongation, including *PRE6* (Hong et al., 2013; Zheng et al., 2017). We extracted genes of highest correlation with these four *GmPRE6s* (weighted value > 0.4) from cyan module and constructed a *GmPRE6s* based submodule. It is worth noting that the genes in this submodule are specifically highly expressed during hypocotyl development (**Supplementary Figure 11**), indicating their important roles in regulating hypocotyl elongation. Interestingly, in an *Arabidopsis thaliana* hypocotyl elongation related module obtained through WGCNA analysis, 27 samples, 33,557 genes, based on RNA-seq results from Burko et al. (2020) (**Supplementary Figure 12**), a homologous submodule was found, i.e., members between the two submodules are homologous genes. Among these genes in the two submodules, PRE6's potential downstream targets, *PIN3* (*At1g70940-Glyma.07g217900*), *PKS4* (*PHYTOCHROME KINASE SUBSTRATE4*, *At5g04190-Glyma.19g123300/Glyma.03g118900*), *IAA29* (*INDOLE-3-ACETIC ACID INDUCIBLE29*, *At4g32280-Glyma.04g066300*), *PRE5* (*At3g28857-Glyma.19g021400*), *SAUR1* (*At4g34770-Glyma.12g033900*) were all reported to be involved in auxin or light regulating hypocotyl elongation (Zheng et al., 2017; Majda and Robert, 2018; Dong et al., 2019; Hu et al., 2021). Interestingly, the 2.5 kb promoter regions of these genes all contain at least one G-box motif (potential binding site of PIFs) (**Supplementary Figure 14E**), indicating they are potential targets of PRE6 related PIFs. A previous study (Hong et al.,

2013) showed that in *Arabidopsis* PRE6 forms non-functional heterodimers with LONG HYPOCOTYL IN FAR-RED1 (HFR1) to activate PIF4 and induce hypocotyl elongation. We note that in the submodule *Glyma.10g138800* (*PHYTOCHROME INTERACTING FACTOR-LIKE15*, *GmPIL15*) is the soybean homolog of *Arabidopsis* HFR1, so it is possible *GmPRE6s* interact with *GmPIL15* to activate *GmPIFs* thus regulating hypocotyl elongation. These results suggest that PRE6 and its downstream genes constitute a conserved regulatory submodule involved in auxin and light regulating hypocotyl elongation (**Figure 5C**).

The four *GmPRE6s* were expressed in the nucleus (**Supplementary Figure 13**). RT-qPCR results showed that their expression was induced by exogenous IAA (**Figure 5D**) and inhibited by light (**Supplementary Figures 14A–D**). Multiple light and auxin responsive elements can be found in the promoter region of *GmPRE6s* (**Supplementary Figure 14E**), which is consistent with the above results. In order to verify whether and how *GmPRE6s* was involved in the regulation of soybean hypocotyl elongation, we ectopically expressed two of *GmPRE6s* (*Glyma.01g044800* and *Glyma.08g298700*) in *Arabidopsis thaliana*, and the result showed that overexpression of *GmPRE6s* induced hypocotyl elongation (**Figure 6**). These results suggest that *GmPRE6s* may participate in hypocotyl elongation pathway regulated by auxin and light factor.

DISCUSSION

Constructing gene regulatory networks through WGCNA has been proved to be an effective approach to mine important phenotype related regulators. By conducting transcriptome analysis at 0, 1 and 5 DAI during soybean hypocotyl development and WGCNA analysis, we established the co-expression regulatory network of soybean hypocotyl elongation (**Figure 1**),



and found a novel cyan module highly related to the phenotype (Figure 2). Most of the genes in this module are enriched in the pathways related to light response, auxin response, and cell wall expansion (Supplementary Figure 6). Thus, the cyan module might be involved in hypocotyl elongation mainly regulated by light and auxin.

Two aspects of evidence show that the cyan module is a reliable candidate of hypocotyl elongation regulation. On the one hand, four out of six mutants of cyan module members' homologous genes in *Arabidopsis thaliana* showed altered hypocotyl elongation phenotype (Supplementary Figure 4); on the other hand, there are 27 genes exist in both the cyan module and previously identified soybean hypocotyl elongation QTLs (Figures 3A,B). Interestingly, we found a conserved submodule that exists in both the cyan module and an *Arabidopsis* module related to hypocotyl elongation, indicating the conservation of hypocotyl elongation mechanism among species (Figure 5C).

In the 27 common genes between the cyan module and previously published QTLs, there are five *SAURs* (Figure 3C). *SAURs* were reported to be involved in the regulation of hypocotyl elongation in *Arabidopsis* and other plants (Wang et al., 2020). Our results show that at least two *SAURs* are direct targets of GmPIF1 and GmPIF3 (Figures 3D,E). PIFs are involved in light and auxin regulating hypocotyl elongation pathways in a variety of plants (Sun et al., 2020). It is not surprising that we found that the expression of *GmSAUR1*, *GmSAUR23*, and *EXPANSINs* were all regulated by auxin and light (Figure 4; Supplementary Figure 9). Based on these results, we constructed the auxin/light-GmPIF1/GmPIF3-GmSAUR1/23-*EXPANSINs* regulatory submodule, in which auxin/light regulates *GmSAUR1* and *GmSAUR23* through GmPIF1 and GmPIF3, and in turn regulates hypocotyl elongation through *EXPANSINs* (Figure 3C; Supplementary Figure 15). Interestingly, in *Arabidopsis thaliana*, light promoted the expression of *SAURs* (*SAUR14*, *SAUR16*, and *SAUR50*), while auxin inhibited (Dong et al., 2019). But in soybean we observed the opposite regulation pattern, i.e., auxin promoted *SAURs*' expression while light inhibited (Figure 4), which may be due to the functional differentiation of *SAUR* genes or the difference between species.

Plant hypocotyl elongation is mainly regulated by transcription factors. There are 83 transcription factors in cyan module, most of which are enriched in auxin and light signal response pathway (Supplementary Figure 10; Figure 5A). We note that four *GmPRE6* genes and several key regulators closely related to them form a regulatory submodule (Figure 5B). In *Arabidopsis*, *PRE6* is involved in the hypocotyl elongation regulated auxin signaling, but the mechanism remains largely elusive (Hong et al., 2013; Zheng et al., 2017). Interestingly, there is a homologous submodule in the *Arabidopsis* regulating hypocotyl elongation (Figure 5C), implying the conservation of relative mechanisms. The expression level of *GmPRE6s* in soybean hypocotyl was induced by exogenous auxin and repressed by light, indicating they may act downstream to regulate hypocotyl elongation (Figure 5D; Supplementary Figures 14A–D). Indeed, ectopic expression of *GmPRE6s* in *Arabidopsis* led to abnormal hypocotyl elongation

(Figure 6A). Interestingly, by analyzing the ChIP-seq data of *Arabidopsis thaliana* (Pfeiffer et al., 2014), we found that PIF3 can bind to the *PRE6* promoter (Supplementary Table 5). The analysis of soybean *GmPRE6* promoters showed that there are auxin/light response elements and G-box elements (Supplementary Figure 14E), indicating that *GmPRE6s* are potential targets of GmPIFs. Therefore, GmPRE6s are promoters of hypocotyl elongation, and light stimulus inhibit their expression to repress hypocotyl elongation, while auxin plays a opposite role (Supplementary Figure 15). GmPRE6s were expressed in the nucleus, but they were reported to be atypical bHLH TFs that lack the basic DNA binding domain, so they may regulate downstream processes by combining with other TFs. In *Arabidopsis*, *PRE6* forms non-functional heterodimers with HFR1 to activate PIFs. In the GmPRE6s submodule, there is a homolog of HFR1, GmPIL15. So GmPRE6s may interact with GmPIL15 to induce the elongation of hypocotyl (Supplementary Figure 15). Future studies could be carried out to clarify the mechanisms.

In conclusion, through WGCNA analysis, we constructed the global regulatory network of soybean hypocotyl elongation for the first time, and identified a key regulatory module. By combining with previous research results, as well as transcription factor analysis, Y1H, mutant and transgenic plant phenotype analysis, we identified two regulatory modules, namely *GmPRE6s-EXPANSINs* submodule and GmPIF1/GmPIF3-GmSAUR1/23-*EXPANSINs* submodule, respectively (Supplementary Figure 15). Our results not only reveal the key regulatory network of hypocotyl elongation, but also narrow down the gene range of hypocotyl elongation regulation, providing valuable gene resources for improving soybean hypocotyl shape and seed vigor through soybean breeding.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: National Center for Biotechnology Information (NCBI) BioProject database under accession number SRR17239198. <https://www.ncbi.nlm.nih.gov/sra/?term=%20SRR17239198>.

AUTHOR CONTRIBUTIONS

MC conceived the study and supervised the whole research. ZS performed experiments and analyzed the data. MC and ZS wrote the manuscript. All authors approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (31970344) to MC.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.837130/full#supplementary-material>

REFERENCES

- Alabadi, D., Gil, J., Blazquez, M. A., and Garcia-Martinez, J. L. (2004). Gibberellins repress photomorphogenesis in darkness. *Plant Physiol.* 134, 1050–1057. doi: 10.1104/pp.103.035451
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37(Web Server issue), W202–208. doi: 10.1093/nar/gkp335
- Bawa, G., Feng, L., Chen, G., Chen, H., Hu, Y., Pu, T., et al. (2020). Gibberellins and auxin regulate soybean hypocotyl elongation under low light and high-temperature interaction. *Physiol. Plant* 170, 345–356. doi: 10.1111/ppl.13158
- Burko, Y., Seluzicki, A., Zander, M., Pedmale, U. V., Ecker, J. R., and Chory, J. (2020). Chimeric activators and repressors define HY5 activity and reveal a light-regulated feedback mechanism. *Plant Cell* 32, 967–983. doi: 10.1105/tpc.19.00772
- Busatto, N., Salvagnin, U., Resentini, F., Quaresimin, S., Navazio, L., Marin, O., et al. (2017). The peach RGF/GLV signaling peptide pCTG134 is involved in a regulatory circuit that sustains auxin and ethylene actions. *Front. Plant Sci.* 8:1711. doi: 10.3389/fpls.2017.01711
- Castelain, M., Le Hir, R., and Bellini, C. (2012). The non-DNA-binding bHLH transcription factor PRE3/bHLH135/ATBS1/TMO7 is involved in the regulation of light signaling pathway in Arabidopsis. *Physiol. Plant* 145, 450–460. doi: 10.1111/j.1399-3054.2012.01600.x
- Castillon, A., Shen, H., and Huq, E. (2009). Blue light induces degradation of the negative regulator phytochrome interacting factor 1 to promote photomorphogenic development of Arabidopsis seedlings. *Genetics* 182, 161–171. doi: 10.1534/genetics.108.099887
- Chen, M., and Penfield, S. (2018). Feedback regulation of COOLAIR expression controls seed dormancy and flowering time. *Science* 360, 1014–1017. doi: 10.1126/science.aar7361
- Cosgrove, D. J. (2016). Catalysts of plant cell wall loosening. *F1000Res* 5. doi: 10.12688/f1000research.7180.1
- de Lucas, M., Daviere, J. M., Rodriguez-Falcon, M., Pontin, M., Iglesias-Pedraz, J. M., Lorrain, S., et al. (2008). A molecular framework for light and gibberellin control of cell elongation. *Nature* 451, 480–484. doi: 10.1038/nature06520
- Deepika, Ankit, Sagar, S., and Singh, A. (2020). Dark-induced hormonal regulation of plant growth and development. *Front. Plant Sci.* 11:581666. doi: 10.3389/fpls.2020.581666
- Dong, J., Sun, N., Yang, J., Deng, Z., Lan, J., Qin, G., et al. (2019). The transcription factors TCP4 and PIF3 antagonistically regulate organ-specific light induction of SAUR genes to modulate cotyledon opening during de-etiolation in Arabidopsis. *Plant Cell* 31, 1155–1170. doi: 10.1105/tpc.18.00803
- Duman, Z., Eliyahu, A., Abu-Abied, M., and Sadot, E. (2020). The contribution of cell wall remodeling and signaling to lateral organs formation. *ISR J. Plant Sci.* 67, 110–127. doi: 10.1163/22238980-20191115
- Fuxman Bass, J. I., Reece-Hoyes, J. S., and Walhout, A. J. (2016). Gene-centered yeast one-hybrid assays. *Cold Spring Harb Protoc.* 2016. doi: 10.1101/pdb.top077669
- Gommers, C. M., Keuskamp, D. H., Buti, S., van Veen, H., Koevoets, I. T., Reinen, E., et al. (2017). Molecular profiles of contrasting shade response strategies in wild plants: differential control of immunity and shoot elongation. *Plant Cell* 29, 331–344. doi: 10.1105/tpc.16.00790
- Hocq, L., Pelloux, J., and Lefebvre, V. (2017). Connecting homogalacturonan-type pectin remodeling to acid growth. *Trends Plant Sci.* 22, 20–29. doi: 10.1016/j.tplants.2016.10.009
- Hong, S. Y., Seo, P. J., Ryu, J. Y., Cho, S. H., Woo, J. C., and Park, C. M. (2013). A competitive peptide inhibitor KIDARI negatively regulates HFR1 by forming nonfunctional heterodimers in Arabidopsis photomorphogenesis. *Mol Cells* 35, 25–31. doi: 10.1007/s10059-013-2159-2
- Hornitschek, P., Kohlen, M. V., Lorrain, S., Rougemont, J., Ljung, K., Lopez-Vidriero, I., et al. (2012). Phytochrome interacting factors 4 and 5 control seedling growth in changing light conditions by directly controlling auxin signaling. *Plant J.* 71, 699–711. doi: 10.1111/j.1365-3113X.2012.05033.x
- Hu, T., Yin, S., Sun, J., Linghu, Y., Ma, J., Pan, J., et al. (2021). Clathrin light chains regulate hypocotyl elongation by affecting the polarization of the auxin transporter PIN3 in Arabidopsis. *J. Integr. Plant Biol.* 63, 1922–1936. doi: 10.1111/jipb.13171
- Hyun, Y., and Lee, I. (2006). KIDARI, encoding a non-DNA Binding bHLH protein, represses light signal transduction in *Arabidopsis thaliana*. *Plant Mol. Biol.* 61, 283–296. doi: 10.1007/s11103-006-0010-2
- Ivakov, A., Flis, A., Apelt, F., Funfgeld, M., Scherer, U., Stitt, M., et al. (2017). Cellulose synthesis and cell expansion are regulated by different mechanisms in growing Arabidopsis hypocotyls. *Plant Cell* 29, 1305–1315. doi: 10.1105/tpc.16.00782
- Jeong, J., and Choi, G. (2013). Phytochrome-interacting factors have both shared and distinct biological roles. *Mol. Cells* 35, 371–380. doi: 10.1007/s10059-013-0135-5
- Jiang, H., Shui, Z., Xu, L., Yang, Y., Li, Y., Yuan, X., et al. (2020). Gibberellins modulate shade-induced soybean hypocotyl elongation downstream of the mutual promotion of auxin and brassinosteroids. *Plant Physiol. Biochem.* 150, 209–221. doi: 10.1016/j.plaphy.2020.02.042
- Jin, H., Li, H., and Zhu, Z. (2021). Experimental procedures for studying skotomorphogenesis in *Arabidopsis thaliana*. *Methods Mol. Biol.* (2297) 2297, 49–60. doi: 10.1007/978-1-0716-1370-2_6
- Josse E. M. and Halliday K. J. (2008). Skotomorphogenesis: the dark side of light signalling. *Curr Biol.* 18, R1144–6. doi: 10.1016/j.cub.2008.10.034
- Kim, J., Song, K., Park, E., Kim, K., Bae, G., and Choi, G. (2016). Epidermal phytochrome B inhibits hypocotyl negative gravitropism non-cell-autonomously. *Plant Cell* 28, 2770–2785. doi: 10.1105/tpc.16.00487
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9:559. doi: 10.1186/1471-2105-9-559
- Lee, K. P., Piskurewicz, U., Tureckova, V., Carat, S., Chappuis, R., Strnad, M., et al. (2012). Spatially and genetically distinct control of seed germination by phytochromes A and B. *Genes Dev.* 26, 1984–1996. doi: 10.1101/gad.194266.112
- Lee, S., Park, K., Lee, H., Park, E., and Boerma, H. (2001). Genetic mapping of QTLs conditioning soybean sprout yield and quality. *Theor. Appl. Genet.* 103, 702–709. doi: 10.1007/s001220100595
- Leivar, P., Monte, E., Oka, Y., Liu, T., Carle, C., Castillon, A., et al. (2008). Multiple phytochrome-interacting bHLH transcription factors repress premature seedling photomorphogenesis in darkness. *Curr. Biol.* 18, 1815–1823. doi: 10.1016/j.cub.2008.10.058
- Leivar, P., Tepperman, J. M., Cohn, M. M., Monte, E., Al-Sady, B., Erickson, E., et al. (2012). Dynamic antagonism between phytochromes and PIF family basic helix-loop-helix factors induces selective reciprocal responses to light and shade in a rapidly responsive transcriptional network in Arabidopsis. *Plant Cell* 24, 1398–1419. doi: 10.1105/tpc.112.095711
- Leivar, P., Tepperman, J. M., Monte, E., Calderon, R. H., Liu, T. L., and Quail, P. H. (2009). Definition of early transcriptional circuitry involved in light-induced reversal of PIF-imposed repression of photomorphogenesis in young Arabidopsis seedlings. *Plant Cell* 21, 3535–3553. doi: 10.1105/tpc.109.070672
- Li, L., Ljung, K., Breton, G., Schmitz, R. J., Pruneda-Paz, J., Cowing-Zitron, C., et al. (2012). Linking photoreceptor excitation to changes in plant architecture. *Genes Dev.* 26, 785–790. doi: 10.1101/gad.187849.112
- Liang, H., Yu, Y., Yang, H., Xu, L., Dong, W., Du, H., et al. (2014). Inheritance and QTL mapping of related root traits in soybean at the seedling stage. *Theor. Appl. Genet.* 127, 2127–2137. doi: 10.1007/s00122-014-2366-z
- Lin, W., Zhou, X., Tang, W., Takahashi, K., Pan, X., Dai, J., et al. (2021). TMK-based cell-surface auxin signalling activates cell-wall acidification. *Nature* 599, 278–282. doi: 10.1038/s41586-021-03976-4
- Liu, S., Zhang, M., Feng, F., and Tian, Z. (2020). Toward a “Green Revolution” for soybean. *Mol. Plant* 13, 688–697. doi: 10.1016/j.molp.2020.03.002
- Liu, Z., Zhang, Y., Liu, R., Hao, H., Wang, Z., and Bi, Y. (2011). Phytochrome interacting factors (PIFs) are essential regulators for sucrose-induced hypocotyl elongation in Arabidopsis. *J. Plant Physiol.* 168, 1771–1779. doi: 10.1016/j.jplph.2011.04.009
- Majda, M., and Robert, S. (2018). The role of auxin in cell wall expansion. *Int. J. Mol. Sci.* 19:951. doi: 10.3390/ijms19040951
- Mara, C. D., Huang, T., and Irish, V. F. (2010). The Arabidopsis floral homeotic proteins APETALA3 and PISTILLATA negatively regulate the BANQUO genes implicated in light signaling. *Plant Cell* 22, 690–702. doi: 10.1105/tpc.109.065946

- Nozue, K., Harmer, S. L., and Maloof, J. N. (2011). Genomic analysis of circadian clock-, light-, and growth-correlated genes reveals PHYTOCHROME-INTERACTING FACTOR5 as a modulator of auxin signaling in Arabidopsis. *Plant Physiol.* 156, 357–372. doi: 10.1104/pp.111.172684
- Oh, E., Zhu, J. Y., and Wang, Z. Y. (2012). Interaction between BZR1 and PIF4 integrates brassinosteroid and environmental responses. *Nat. Cell Biol.* 14, 802–809. doi: 10.1038/ncb2545
- Paque, S., Mouille, G., Grandont, L., Alabadi, D., Gaertner, C., Goyallon, A., et al. (2014). AUXIN BINDING PROTEIN1 links cell wall remodeling, auxin signaling, and cell expansion in Arabidopsis. *Plant Cell* 26, 280–295. doi: 10.1105/tpc.113.120048
- Pelletier, S., Van Orden, J., Wolf, S., Vissenberg, K., Delacourt, J., Ndong, Y. A., et al. (2010). A role for pectin de-methylesterification in a developmentally regulated growth acceleration in dark-grown Arabidopsis hypocotyls. *New Phytol.* 188, 726–739. doi: 10.1111/j.1469-8137.2010.03409.x
- Pfeiffer, A., Shi, H., Tepperman, J. M., Zhang, Y., and Quail, P. H. (2014). Combinatorial complexity in a transcriptionally centered signaling hub in Arabidopsis. *Mol. Plant* 7, 1598–1618. doi: 10.1093/mp/ssu087
- Qiu, Y., Pasorek, E. K., Reddy, A. K., Nagatani, A., Ma, W., Chory, J., et al. (2017). Mechanism of early light signaling by the carboxy-terminal output module of Arabidopsis phytochrome B. *Nat. Commun.* 8:1905. doi: 10.1038/s41467-017-02062-6
- Reed, J. W., Wu, M. F., Reeves, P. H., Hodgson, C., Yadav, V., Hayes, S., et al. (2018). Three auxin response factors promote hypocotyl elongation. *Plant Physiol.* 178, 864–875. doi: 10.1104/pp.18.00718
- Sassi, M., Wang, J., Ruberti, I., Vernoux, T., and Xu, J. (2013). Shedding light on auxin movement: light-regulation of polar auxin transport in the photocontrol of plant development. *Plant Signal. Behav.* 8:e23355. doi: 10.4161/psb.23355
- Seo, M., Nambara, E., Choi, G., and Yamaguchi, S. (2009). Interaction of light and hormone signals in germinating seeds. *Plant Mol. Biol.* 69, 463–472. doi: 10.1007/s11103-008-9429-y
- 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
- Shen, H., Zhu, L., Castillon, A., Majee, M., Downie, B., and Huq, E. (2008). Light-induced phosphorylation and degradation of the negative regulator PHYTOCHROME-INTERACTING FACTOR1 from Arabidopsis depend upon its direct physical interactions with photoactivated phytochromes. *Plant Cell* 20, 1586–1602. doi: 10.1105/tpc.108.060020
- Shin, J., Kim, K., Kang, H., Zulfugarov, I. S., Bae, G., Lee, C. H., et al. (2009). Phytochromes promote seedling light responses by inhibiting four negatively-acting phytochrome-interacting factors. *Proc. Natl. Acad. Sci. U. S. A.* 106, 7660–7665. doi: 10.1073/pnas.0812219106
- Srivastava, A. K., Senapati, D., Srivastava, A., Chakraborty, M., Gangappa, S. N., and Chattopadhyay, S. (2015). Short Hypocotyl in White Light1 interacts with Elongated Hypocotyl5 (HY5) and Constitutive Photomorphogenic1 (COP1) and promotes COP1-mediated degradation of HY5 during Arabidopsis seedling development. *Plant Physiol.* 169, 2922–2934. doi: 10.1104/pp.15.01184
- Sun, W., Han, H., Deng, L., Sun, C., Xu, Y., Lin, L., et al. (2020). Mediator subunit MED25 physically interacts with PHYTOCHROME INTERACTING FACTOR4 to regulate shade-induced hypocotyl elongation in tomato. *Plant Physiol.* 184, 1549–1562. doi: 10.1104/pp.20.00587
- Wang, H., and Wang, H. (2015). Phytochrome signaling: time to tighten up the loose ends. *Mol. Plant* 8, 540–551. doi: 10.1016/j.molp.2014.11.021
- Wang, L., Yu, P., Lyu, J., Hu, Y., Han, C., Bai, M. Y., et al. (2021). BZR1 physically interacts with SPL9 to regulate the vegetative phase change and cell elongation in Arabidopsis. *Int. J. Mol. Sci.* 22:10415. doi: 10.3390/ijms221910415
- Wang, X., Yu, R., Wang, J., Lin, Z., Han, X., Deng, Z., et al. (2020). The asymmetric expression of SAUR genes mediated by ARF7/19 promotes the gravitropism and phototropism of plant hypocotyls. *Cell Rep.* 31:107529. doi: 10.1016/j.celrep.2020.107529
- Xin, X., Lei, L., Zheng, Y., Zhang, T., Pingali, S. V., O'Neill, H., et al. (2020). Cellulose synthase interactive1- and microtubule-dependent cell wall architecture is required for acid growth in Arabidopsis hypocotyls. *J. Exp. Bot.* 71, 2982–2994. doi: 10.1093/jxb/eraa063
- Yi, R., Yan, J., and Xie, D. (2020). Light promotes jasmonate biosynthesis to regulate photomorphogenesis in Arabidopsis. *Sci. China Life Sci.* 63, 943–952. doi: 10.1007/s11427-019-1584-4
- Yu, G., Wang, L. G., Han, Y., and He, Q. Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yu, Y., Wang, J., Zhang, Z., Quan, R., Zhang, H., Deng, X. W., et al. (2013). Ethylene promotes hypocotyl growth and HY5 degradation by enhancing the movement of COP1 to the nucleus in the light. *PLoS Genet.* 9:e1004025. doi: 10.1371/journal.pgen.1004025
- Zaidi, S. S., Naqvi, R. Z., Asif, M., Strickler, S., Shakir, S., Shafiq, M., et al. (2020). Molecular insight into cotton leaf curl geminivirus disease resistance in cultivated cotton (*Gossypium hirsutum*). *Plant Biotechnol. J.* 18, 691–706. doi: 10.1111/pbi.13236
- Zhang, W., Tang, Y., Hu, Y., Yang, Y., Cai, J., Liu, H., et al. (2021). Arabidopsis NF-YCs play dual roles in repressing brassinosteroid biosynthesis and signaling during light-regulated hypocotyl elongation. *Plant Cell* 33, 2360–2374. doi: 10.1093/plcell/koab112
- Zhang, Y., Mayba, O., Pfeiffer, A., Shi, H., Tepperman, J. M., Speed, T. P., et al. (2013). A quartet of PIF bHLH factors provides a transcriptionally centered signaling hub that regulates seedling morphogenesis through differential expression-patterning of shared target genes in Arabidopsis. *PLoS Genet.* 9:e1003244. doi: 10.1371/journal.pgen.1003244
- Zhao, Q., Yuan, S., Wang, X., Zhang, Y., Zhu, H., and Lu, C. (2008). Restoration of mature etiolated cucumber hypocotyl cell wall susceptibility to expansin by pretreatment with fungal pectinases and EGTA *in vitro*. *Plant Physiol.* 147, 1874–1885. doi: 10.1104/pp.108.116962
- Zheng, K., Wang, Y., Zhang, N., Jia, Q., Wang, X., Hou, C., et al. (2017). Involvement of PACLOBUTRAZOL RESISTANCE6/KIDARI, an atypical bHLH transcription factor, in auxin responses in Arabidopsis. *Front. Plant Sci.* 8:1813. doi: 10.3389/fpls.2017.01813
- Zhong, M., Zeng, B., Tang, D., Yang, J., Qu, L., Yan, J., et al. (2021). The blue light receptor CRY1 interacts with GID1 and DELLA proteins to repress GA signaling during photomorphogenesis in Arabidopsis. *Mol. Plant* 14, 1328–1342. doi: 10.1016/j.molp.2021.05.011
- Zhong, S., Shi, H., Xue, C., Wang, L., Xi, Y., Li, J., et al. (2012). A molecular framework of light-controlled phytohormone action in Arabidopsis. *Curr. Biol.* 22, 1530–1535. doi: 10.1016/j.cub.2012.06.039

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 Shen and Chen. 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.



CRISPR/Cas9-Mediated Targeted Mutagenesis of *GmUGT* Enhanced Soybean Resistance Against Leaf-Chewing Insects Through Flavonoids Biosynthesis

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Jia-He Wu,
Institute of Microbiology, Chinese
Academy of Sciences (CAS), China
Xiangdong Yang,
Jilin Academy of Agricultural Sciences
(CAAS), China

*Correspondence:

Dong Cao
caodong@caas.cn
Xia Li
xli@mail.hzau.edu.cn
Yongqing Jiao
yqjiao@126.com

† These authors have contributed
equally to this work

Specialty section:

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

Received: 27 October 2021

Accepted: 07 January 2022

Published: 22 February 2022

Citation:

Zhang Y, Guo W, Chen L, Shen X,
Yang H, Fang Y, Ouyang W, Mai S,
Chen H, Chen S, Hao Q, Yuan S,
Zhang C, Huang Y, Shan Z, Yang Z,
Qiu D, Zhou X, Cao D, Li X and Jiao Y
(2022) CRISPR/Cas9-Mediated
Targeted Mutagenesis of *GmUGT*
Enhanced Soybean Resistance
Against Leaf-Chewing Insects
Through Flavonoids Biosynthesis.
Front. Plant Sci. 13:802716.
doi: 10.3389/fpls.2022.802716

Yongxing Zhang^{1,2†}, Wei Guo^{1†}, Limiao Chen¹, Xinjie Shen¹, Hongli Yang¹, Yisheng Fang¹, Wenqi Ouyang¹, Sihua Mai¹, Haifeng Chen¹, Shuilian Chen¹, Qingnan Hao¹, Songli Yuan¹, Chanjuan Zhang¹, Yi Huang¹, Zhihui Shan¹, Zhonglu Yang¹, Dezhen Qiu¹, Xinan Zhou¹, Dong Cao^{1*}, Xia Li^{2*} and Yongqing Jiao^{1,3*}

¹ Key Laboratory of Biology and Genetic Improvement of Oil Crops, Ministry of Agriculture and Rural Affairs, Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, Wuhan, China, ² National Key Laboratory of Crop Genetic Improvement, College of Plant Science and Technology, Huazhong Agricultural University, Wuhan, China, ³ Collaborative Innovation Center of Henan Grain Crops, College of Agronomy, Henan Agricultural University, Zhengzhou, China

Leaf-chewing insects are important pests that cause yield loss and reduce seed quality in soybeans (*Glycine max*). Breeding soybean varieties that are resistant to leaf-chewing insects can minimize the need for insecticide use and reduce yield loss. The marker gene for QTL-M, *Glyma.07g110300* (LOC100775351) that encodes a UDP-glycosyltransferase (UGT) is the major determinant of resistance against leaf-chewing insects in soybean; it exhibits a loss of function in insect-resistant soybean germplasms. In this study, *Agrobacterium*-mediated transformation introduced the CRISPR/Cas9 expression vector into the soybean cultivar Tianlong No. 1 to generate *Glyma.07g110300*-gene mutants. We obtained two novel types of mutations, a 33-bp deletion and a single-bp insertion in the *GmUGT* coding region, which resulted in an enhanced resistance to *Helicoverpa armigera* and *Spodoptera litura*. Additionally, overexpressing *GmUGT* produced soybean varieties that were more sensitive to *H. armigera* and *S. litura*. Both mutant and overexpressing lines exhibited no obvious phenotypic changes. The difference in metabolites and gene expression suggested that *GmUGT* is involved in imparting resistance to leaf-chewing insects by altering the flavonoid content and expression patterns of genes related to flavonoid biosynthesis and defense. Furthermore, ectopic expression of the *GmUGT* gene in the *ugt72b1* mutant of *Arabidopsis* substantially rescued the phenotype of *H. armigera* resistance in the *atugt72b1* mutant. Our study presents a strategy for increasing resistance against leaf-chewing insects in soybean through CRISPR/Cas9-mediated targeted mutagenesis of the *UGT* genes.

Keywords: CRISPR/Cas9, soybean, UDP-glycosyltransferase, *Helicoverpa armigera* Hübner, *Spodoptera litura* Fabricius

INTRODUCTION

Soybean (*Glycine max*) is an important industrial crop that provides edible oil, vegetable protein, and active compounds, such as flavonoids (Arifin et al., 2021). Leaf-chewing insects, such as *Helicoverpa zea*, *Helicoverpa armigera* Hübner, and *Spodoptera litura* Fabricius, are important insect pests that cause a loss in yield and a decline in the quality of soybeans (Otuka et al., 2020; Haile et al., 2021). Farmers mainly depend on chemical pesticides to control this threat. However, the long-term use of chemical pesticides leads to the development of drug resistance in the target insects, contaminates land and water resources, and increases costs of environmental management (Elba et al., 2014; Bondori et al., 2018). Breeding soybean varieties that are resistant to leaf-chewing insects can minimize the need for insecticides, improve soybean yields, and reduce concerns stemming from pesticide use (Ortega et al., 2016).

Transgenic breeding offers key breakthroughs in plant breeding, and the cultivation of transgenic crops has appreciably increased world agricultural productivity in the past two decades (Raman, 2017). Genetically modified cultivars expressing insecticidal proteins have become vital to the effective management of leaf-chewing insect pests. Major insecticidal proteins include the crystal protein (CRY), the vegetative insecticidal protein (VIP), and the protease inhibitor (PI), whose genes have been used to develop insect-resistant genetically modified cultivars (Saikhedkar et al., 2019; Kumar et al., 2020). Many insect-resistant transgenic crops now exist; there are six commercialized soybean cultivars with various insect-resistance (IR) genes (Ichim, 2019; Kumar et al., 2020). This genetic trait provides an effective control strategy for defoliators, including *H. armigera*. However, transgenic crops that have foreign DNA randomly integrated into their genomes have always been controversial, owing to the durability of resistance, environmental safety, and potential adverse health effects on consumers, limiting such transgenic breeding (Domingo and Bordonaba, 2011; Raman, 2017; Schiemann et al., 2019; Pozebon et al., 2020; Gao, 2021).

In addition to the deployment of genetically modified soybean varieties that overexpress one or more insecticidal protein genes, other soybean varieties have been screened for native resistance to leaf-chewing insects, such as *H. armigera* (Haile et al., 2021). Zhu et al. (2006) fine-mapped a major quantitative trait locus (QTL-M) for insect resistance in PI 229358 and developed Benning near-isogenic lines (NILs) resistant to *H. zea*. Similarly, several studies have reported insect-resistance QTLs against *S. litura* (Kim et al., 2013, 2015; Liu et al., 2016). Ghione et al. (2021) found four markers associated with resistance genes against stink bugs using the association mapping strategy. To date, several insect-resistance QTLs, specific for leaf-chewing insects, have been identified in the soybean germplasm (Rector et al., 2000; Zhu et al., 2008; Wang et al., 2015; Ortega et al., 2016; Sabljic et al., 2020; Ghione et al., 2021; Lucini et al., 2021). Interestingly, Ortega et al. (2017) reported that the *Glyma.07g110300* marker (encoding a UDP-glycosyltransferase) in QTL-M was a functional single-nucleotide polymorphism (SNP) that could be used for

the marker-assisted selection of insect-resistance QTLs and *Glyma.07g110300* exhibits a loss of function in insect-resistant soybean germplasms. Native resistance traits have been reported against *H. zea*, but information about the *GmUGT* gene and its roles in resistance to *H. armigera* remains unclear.

The CRISPR/Cas9 (clustered regularly interspaced short palindromic repeat-associated endonucleases) is an efficient tool for genetic manipulation; it uses simple RNA-guided DNA recognition and binding for sequence-specific nucleic acid cleavage. The CRISPR/Cas9-based genome-editing can help produce crop varieties, similar to non-transgenic crops, without adding any foreign DNA to the genome, and thus, has become a powerful tool to advance crop breeding (Knott and Jennifer, 2018; Chen et al., 2019; Schiemann et al., 2019; Gao, 2021). This technology has been widely applied in soybeans (Bao et al., 2019; Cheng et al., 2019; Han et al., 2019; Cai et al., 2020). Therefore, CRISPR/Cas9-mediated construction of insect-resistant germplasms to reduce the dependence on pesticides has the potential to become a simple and feasible breeding method for soybeans in the future.

In this study, we employed the CRISPR/Cas9 system to create *GmUGT*-gene knockout mutants and analyzed the change in resistance of the mutant to leaf-chewing insects. The *H. armigera* and *S. litura* larval feeding assay demonstrated that the mutagenesis of *GmUGT*-improved soybean resistance against leaf-chewing insects by altering the biosynthesis pathway of flavonoids in soybeans. Furthermore, the ectopic expression of the *GmUGT* gene substantially rescued the phenotype of resistance against *H. armigera* in the *atugt72b1* mutant. Findings from this study present a strategy to increase the resistance to leaf-chewing insects in soybean and other crops through CRISPR/Cas9-mediated targeted mutagenesis of *UGT* genes.

MATERIALS AND METHODS

Plant Materials and Growth Conditions

Soybean cultivar Tianlong No. 1 (WT), as well as transgenic varieties generated by the CRISPR/Cas9-mediated mutagenesis of *GmUGT* gene and *GmUGT*-overexpression, were grown in pots (height × top diameter = 18.5 cm × 18.5 cm). Three plants per pot were grown in an artificial climate chamber under 14-h light/10-h dark photoperiod conditions at 26°C. *Arabidopsis thaliana* Columbia-0 (Col-0) and the T-DNA insertion mutant, *atugt72b1*, were obtained from the Arabidopsis Biological Resource Center (ABRC). The Col-0, *atugt72b1*, and transgenic *Arabidopsis* lines were grown in pots (height × top diameter = 10 × 8 cm) with four plants per pot in an artificial climate chamber under 8-h light/16-h dark photoperiod conditions at 22°C.

Vector Constructs and Plants Transformation

To edit the *GmUGT* gene, the pSC1-CRISPR/Cas9P_{GmUbi3}-BK plasmids were constructed, following the method of Du et al. (2016). Briefly, the target adaptor was designed using

the CRISPR-P¹ web tool; it was then integrated into the single guide RNA (sgRNA) expression cassettes driven by the *GmU6* promoter, which was subsequently built into the pSC1-CRISPR/Cas9P_{GmUbi3}-BK vector. The full-length coding sequence (CDS) of *GmUGT* was cloned from the Tianlong No. 1 complementary DNA (cDNA) library. The *GmUGT* CDS was subcloned into the pB2GW7.0 binary vector (Shen et al., 2018) through the LR recombinase reaction (Invitrogen, Carlsbad, CA, United States). The constructs, *P35S-GmUGT*-pB2GW7.0 and *P_{GmU6}-sgRNA*-pSC1-CRISPR/Cas9P_{GmUbi3}-BK, were introduced into *Agrobacterium tumefaciens* EHA105 by electroporation and then transformed into the WT as previously described (Chen et al., 2021).

For *Arabidopsis* transformation, the *GmUGT* gene was subcloned into pFGC5941 (Lei et al., 2021) driven by the *ATUGT72B1* promoter using the ClonExpress II One Step Cloning Kit (Vazyme, Nanjing, China); the clone, thus, prepared was introduced into the *Arabidopsis atugt72b1* mutant with EHA105 by the floral dip method (Clough and Bent, 1998).

Identification of Transformants and Mutations

For PCR, genomic DNA was extracted from the leaves of the WT plants and the transgenic varieties generated by the CRISPR/Cas9-mediated mutagenesis of *GmUGT* and *GmUGT*-overexpression. Transgenic plants were tested with a Transgenic Plant Bar Gene Rapid Detection Kit (Institute of Oil Crops, Chinese Academy of Agricultural Sciences, Wuhan, China).

The PCR products were separated by electrophoresis on 1% agarose in 1 × Tris-acetate-EDTA (TAE) buffer to screen the *GmUGT* mutants. The purified DNA fragments with *Bar* gene detection primers were sequenced and analyzed. To identify mutations in the regenerated plants, amplified DNA surrounding the target regions of the sgRNA was sequenced and analyzed. The successfully edited types were identified by sequence alignment with the WT DNA sequence. The PCR identified *GmUGT*-overexpressing transgenic soybean plants with vector detection primers. The real-time quantitative PCR (qRT-PCR) was used to determine the expression levels of *GmUGT* in overexpressing transgenic lines. The *GmSKIP16* gene was used as the internal control in qRT-PCR analysis. Rescued lines of *Arabidopsis* were selected on the soil by spraying Basta at 1:1000 concentrations (Bayer CropScience, Monheim, Germany). The qRT-PCR was used to determine the expression levels of *GmUGT* in transgenic *Arabidopsis* lines. The *AtACTIN2* gene was used as the internal control in qRT-PCR analysis.

Detection of Off-Target Sites

To consider the off-target effects, the potential off-target sites were predicted by the website CCTop². The overlapping sequence between the similar position of the off-target site gene (*Glyma.09G245300*) and the sgRNA was 16-bp long. This gene was checked by PCR amplification and sequencing. Off-target site detection primers are listed in **Supplementary Table 1**.

¹<http://cbi.hzau.edu.cn/crispr/>

²<https://cctop.cos.uni-heidelberg.de:8043>

Analysis of Insect Resistance in Mutant and Transgenic Lines

For the larval feeding assay, soybean plants were grown until they had five fully unfolded leaves. *Arabidopsis* plants were grown until they had 10 fully unfolded leaves. The *H. armigera* and *S. litura* larvae obtained from Huazhong Agricultural University were hatched at 28°C from a single egg mass. To analyze insect resistance in the WT plant and the transgenic varieties from CRISPR/Cas9-mediated mutagenesis of the *GmUGT* gene and *GmUGT* overexpression, leaves were collected from the same positions. Four small rounded blades (each with a diameter of 4 cm) were collected from each leaf with a hole puncher. The leaves were placed in a Petri dish (90 × 15 mm) containing two pieces of soaked filter paper. A second instar larva was placed on each leaf. Three days after the treatment, representative photos were taken, and the leaf area of soybean varieties was measured using a leaf area meter. After feeding for 7 days, the weight per larvae was recorded. Two knockout mutants (*ko-3* and *ko-5*), three overexpressing transgenic lines (OX-1, OX-2, and OX-44), and the WT plants underwent fifty biological replicates. For the analysis of insect resistance in *Arabidopsis*, two whole plants were used as one sample. Plants were wrapped in moistened absorbent cotton and placed in 500 mL round disposable plastic boxes with 100 freshly hatched larvae. Three days after treatment, representative photos were taken. After feeding for 7 days, the weight per larvae and the number of surviving larvae were recorded. Three transgenic lines of soybean (RE-1, RE-2, and RE-3), *atugt72b1*, and Col-0 were tested in three biological replicates.

Expression Pattern Analysis of *GmUGT*

The tissues of the WT plants were tested at different stages to analyze the expression pattern of *GmUGT* in soybean. Cotyledons, roots, simple leaves, trifoliate leaves, and stems were harvested from the WT at the V2 stage, flowers were harvested at the R2 stage, and pods at the R5 stage for RNA purification. The harvested samples were stored at −80°C until total RNA extraction was performed for gene expression analyses.

For the induced expression pattern analysis, 10 *H. armigera* larvae were placed on the first and second trifoliate leaves of the WT at the V4 stage. The leaves of the control and treated plants were excised at nine sampling times (0, 1, 2, 3, 6, 12, 24, 36, 48, or 72 h) after being attacked by *H. armigera* to identify the induced resistance. For mechanical damage treatment, the damage was simulated using a needle with 48 spines per leaf. A mix of damaged and undamaged leaves was collected. The *H. armigera* attack and mechanical damage treatments were performed in biological triplicates.

RNA Isolation and Real-Time Quantitative PCR Analysis

Total RNA was extracted from plant tissues using TRIzol reagent (Invitrogen, Carlsbad, CA, United States). The RNA concentration was measured using NanoDrop One (Thermo Scientific, Waltham, MA, United States) for qRT-PCR and RNA sequence (RNA-seq) assays. One microgram of total RNA per sample was used to synthesize the first-strand of cDNA

using the TransScript® First-Strain cDNA Synthesis SuperMix according to the manufacturer's instructions (TransGen Biotech, Beijing, China). A total volume of 20 μ L was set up for the qRT-PCR that contained 0.5 μ M forward and reverse primers, 10 μ L of the SYBR premix, and 20 ng of the cDNA. The assay was performed in a QuantStudio™ 5 Real-Time PCR Instrument (Thermo Scientific, Waltham, MA, United States) under the following cycle conditions: 95°C for 30 s, 40 cycles of denaturation at 95°C for 5 s, annealing at 60°C for 30 s, and extension at 72°C for 30 s with simultaneous fluorescence measurement and a melting curve at 55–95°C (0.5°C increments with each cycle). Three biological replicates were assayed. The *GmSKIP16* (*Glyma.12g05510*), *AtACTIN2* (*At3G18780*), and *HaActin* (*GU182917*) were used as endogenous controls for experiments involving nucleic acids from *Arabidopsis*, soybean, and *H. armigera*, respectively. The expression levels of the target genes were quantified using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). Unless otherwise indicated, the wild-type untreated samples were set to a relative value of 1. All the primers used for qRT-PCR are listed in **Supplementary Table 1**.

Subcellular Localization of GmUGT-YFP Fusion Proteins

The full-length CDS of *GmUGT* was amplified with two primers, *GmUGT*-CDS-F/R (**Supplementary Table 1**). The PCR product was subcloned into the pEarlyGate 101 vector (Deguchi et al., 2020) under the control of the CaMV35S promoter using the ClonExpress II One Step Cloning Kit (Vazyme, Nanjing, China). This yielded a pEarlyGate101-*GmUGT*-YFP construct, which was transformed into *Agrobacterium* strain GV31011 by electroporation. The *Nicotiana benthamiana* plants were grown in a plant growth chamber at 26°C under a 16 h/8 h (light/dark) regimen until they reached heights of approximately 10–20 cm. The plantlets were then infiltrated with *Agrobacterium* strain GV3101 containing pEarlyGate101-*GmUGT*-YFP. Infiltration was performed as described by Bai et al. (2013). The YFP alone was used as the control, and FIBRILLARIN2-mCherry (FIB2-mCherry) was used as a nucleolar marker (Missbach et al., 2013). The agroinfiltrated leaves were photographed 72 h after infiltration. The fluorescing fusion protein was visualized by confocal microscopy with excitation at 520 nm for YFP and 570–620 nm for FIB2-mCherry.

Metabolomics Analysis

For the metabolomics analysis, samples were isolated from the WT and the *ko-3* leaves; these leaves included unattacked and attacked samples by *H. armigera* for 36 h and were taken from the same place on the plants. Nine fully expanded leaves from nine independent plants were pooled as one composite biological replicate. Three biological replicates were performed. Biological samples were extracted, and the sample extracts were analyzed using a ultra-performance liquid chromatography with electrospray ionization tandem mass spectrometry (UPLC-ESI-MS/MS) system as described previously (Chen et al., 2013). Significantly regulated metabolites between groups were determined by variable importance in projection (VIP) ≥ 1 and absolute log₂ fold change (log₂FC) ≥ 1 . The VIP values were extracted from the Orthogonal Partial Least Squares

Discrimination Analysis (OPLS-DA) results, which contained scores and permutation plots generated using the R package, MetaboAnalystR 3.0³. The data were log-transformed (log₂) and mean-centered before OPLS-DA. Identified metabolites were annotated using the KEGG Compound database⁴, and annotated metabolites were then mapped to the Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database⁵. Pathways with significantly regulated metabolites were mapped and then fed into metabolite set enrichment analysis (MSEA). Their significance was determined by the *p* values of the hypergeometric tests (Thevenot et al., 2015; Zhang X. L. et al., 2017).

RNA-Seq and Bioinformatics Analysis

Six fully expanded leaves from three independent WT and *ko-3* plants after *H. armigera* larval attack and unattack were pooled as a single biological replicate. The three biological replicates were used for total RNA extraction followed by RNA-seq analysis. The RNA-seq experiment and preliminary data analysis were carried out by the Beijing Biomarker Technology Corporation (Biomarker, Beijing, China). The total RNA was extracted, and sequencing libraries were generated as previously described (Kou et al., 2021). Clean reads were mapped to the soybean reference genome (*G. max* Wm82.a2.v1) downloaded from Phytozome using SOAP aligner/SOAP2. Differentially expressed genes (DEGs) relative to control samples were identified with the absolute value of |log₂ FC| ≥ 1 and false discovery rate (FDR) < 0.5 as the threshold. The KEGG pathway enrichment of the DEGs was generated using R using enrichment factors, Q-values, and the number of enriched genes in this pathway (Chen et al., 2020).

Statistical Analysis

All data were collected and analyzed using Microsoft Office Excel 2010. All data analyses were performed by Student's *t*-test to compare the means of two samples (or treatments) to determine significant differences. Unless otherwise specified, experiments were conducted in triplicates. The values are given as mean \pm SD. The consistency between RNA-seq and qRT-PCR analyses was evaluated by the Pearson correlation method based on log₂FC. Pearson correlation analyses were performed using GraphPad Prism 8.0 software (GraphPad Software, San Diego, CA, United States).

RESULTS

CRISPR/Cas9-Mediated Mutagenesis of *GmUGT* in Soybean Enhanced Resistance to *Helicoverpa armigera* and *Spodoptera litura*

To better understand the *GmUGT* gene and its function in soybean resistance against leaf-chewing insects, *Agrobacterium*-mediated transformation was used to introduce the CRISPR/Cas9

³<https://github.com/xia-lab/MetaboAnalystR>

⁴<http://www.kegg.jp/kegg/compound/>

⁵<http://www.kegg.jp/kegg/pathway.html>

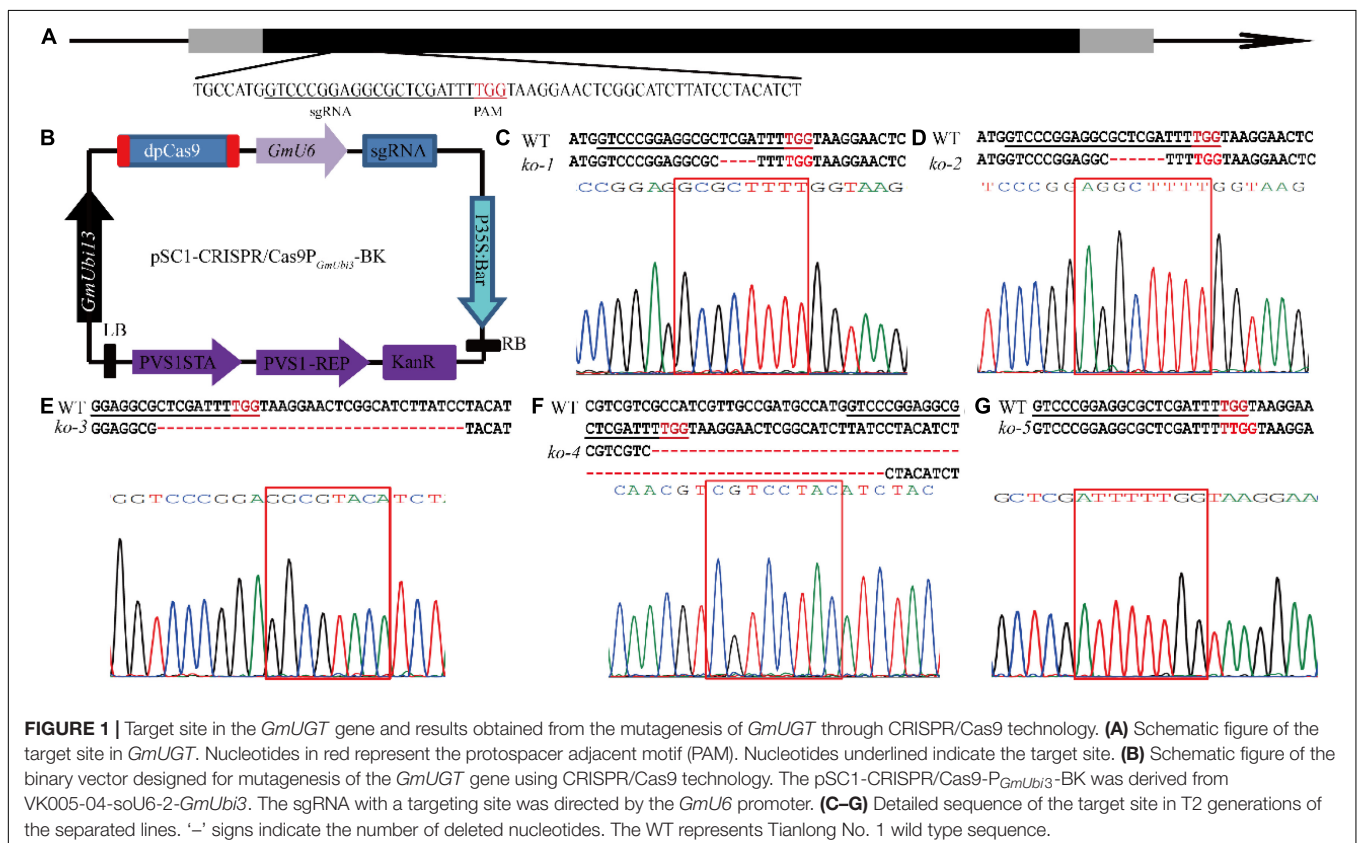
expression vector into the soybean cultivar Tianlong No. 1 (wild type, WT). This generated mutants of the *Glyma.07g110300* gene with a guide RNA targeting *GmUGT* (Figures 1A,B). We obtained four T0 transgenic lines with insertions of the *Bar* gene (Supplementary Figures 1A,B). Several T1 generations were sequenced to identify the CRISPR/Cas9-induced mutant lines, and five edited mutants were identified (Figures 1C–G and Supplementary Figure 1C). The *ko-1*, *ko-2*, *ko-3*, and *ko-4* mutants had 4-bp, 6-bp, 33-bp, and 65-bp deletions, respectively, while *ko-5* had a single-bp insertion in the target site, resulting in altered amino acid sequence or frameshift mutations in the protein encoded by *GmUGT* (Figures 1C–G and Supplementary Figure 2A). Due to the possibility of off-targeting, potential off-target sites predicted by the website CCTop⁶ were also sequenced to detect all cleavages in the mutants (Supplementary Figure 2B). We did not observe a significant difference in plant growth and development between the five homozygous mutants and the WT plants (Supplementary Figure 1C). The ‘transgene-clean’ homozygous *ko-3* and *ko-5* mutants were used for further analysis (Supplementary Figures 1D–F).

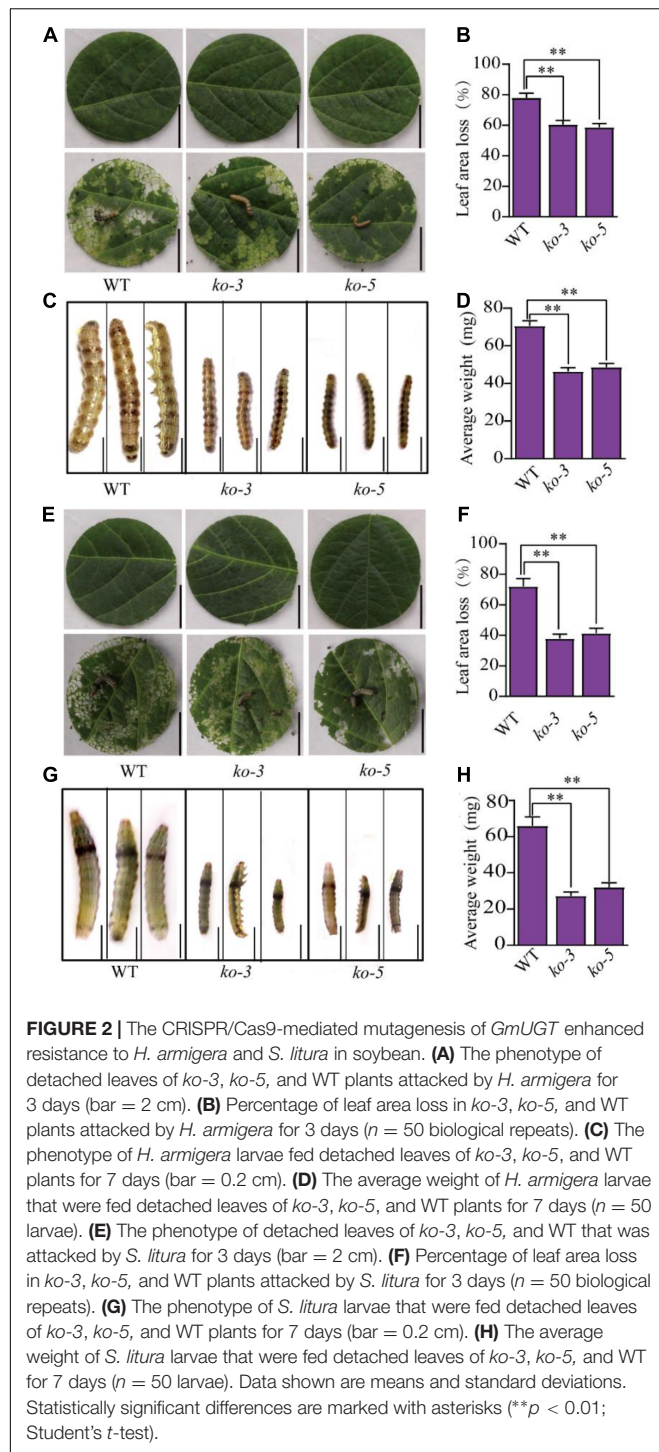
To test the resistance of the *GmUGT* mutants to leaf-chewing insects, we fed the larvae of *H. armigera* and *S. litura* detached leaves of the *ko-3* and *ko-5* mutants and the WT plants. Figure 2 shows that both the *ko-3* and *ko-5* mutants exhibited resistance to *H. armigera* and *S. litura*. The leaf area loss in the *ko-3* and *ko-5* mutants attacked by *H. armigera* was significantly

less than that in the WT plants (Figures 2A,B). The average weight of *H. armigera* fed with the leaves of *ko-3* and *ko-5* mutants was also significantly less than that fed with leaves of the WT plants (Figures 2C,D). Similar results were observed when the larvae of *S. litura* were fed detached leaves of the *ko-3* and *ko-5* mutants and the WT plants (Figures 2E–H). These results demonstrated that the CRISPR/Cas9-mediated targeted mutagenesis of *GmUGT* increased resistance against *H. armigera* and *S. litura* in soybeans.

Soybean Varieties Overexpressing *GmUGT* Were More Sensitive to *H. armigera* and *S. litura*

Because the *GmUGT* mutants (*ko-3* and *ko-5*) showed enhanced resistance to leaf-chewing insects, we wondered whether the overexpression of *GmUGT* would be more sensitive to *H. armigera* and *S. litura*. Four transgenic lines that overexpressed the *GmUGT* gene were generated (OX-1, OX-2, OX32, and OX-44) (Supplementary Figure 3); three of these (OX-1, OX-2, and OX-44) were used to evaluate resistance against *H. armigera* and *S. litura*. As expected, all three *GmUGT*-overexpressing lines were more sensitive to *H. armigera* and *S. litura* in soybean (Figure 3). The loss in leaf area in *GmUGT*-overexpressing lines attacked by *H. armigera* and *S. litura* was more significant than in the WT plants. Furthermore, the average weights of larvae fed *GmUGT*-overexpressing leaves were higher than those that were fed WT

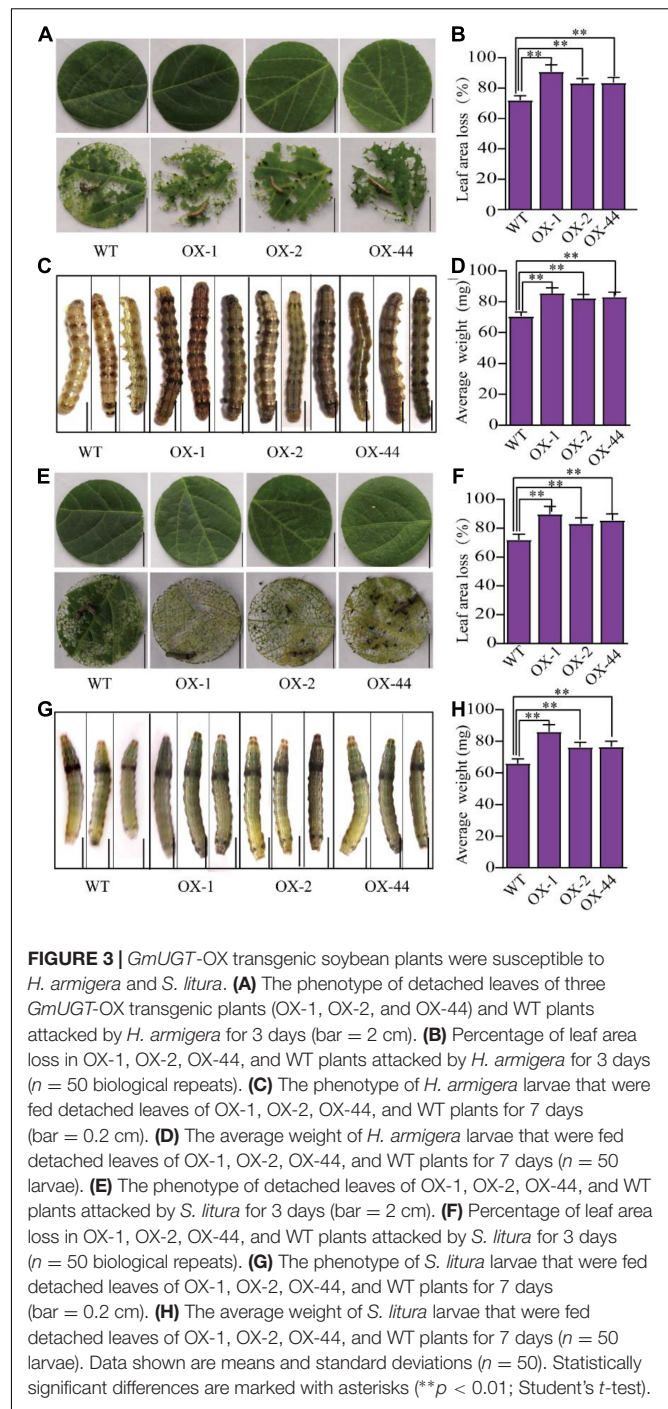




leaves (Figure 3). These results suggest that the *GmUGT* gene compromises the resistance of soybeans to leaf-chewing insects.

Characterization of the *GmUGT* Gene in Soybeans

The tissue expression pattern showed that the *GmUGT* gene was ubiquitously expressed in all tissues. The expression levels in the



flower, simple leaf, trifoliate leaves, and pod were higher than those in other tissues (Figure 4A). Leaves are the favorite food of *H. armigera* and *S. litura* larvae. Additionally, Aljbory and Chen (2018) reported that insect attacks can stimulate plant resistance by mechanical damage and/or the application of an elicitor. Thus, we examined the expression patterns of the *GmUGT* gene in the leaves of WT plants after exposure to *H. armigera* attacks and mechanical damage at different times. The transcription of *GmUGT* was induced by *H. armigera* 1 h after the attack,

followed by a gradual increase for 36 h and eventually a decrease after 48 h; contrarily, the expression levels of *GmUGT* remained unchanged after mechanical damage (Figure 4B). These results indicate that *GmUGT* is expressed in response to the attack of leaf-chewing insects rather than mechanical damage alone in soybean. Phylogenetic analysis showed that the GmUGT protein shared high sequence homology with ATUGT72B1 in *Arabidopsis* (Figure 4C), which catalyzes the glucose conjugation of monolignols (Lin et al., 2016). Previous works have reported that genes belonging to the UGT family function as glycosyltransferase in the cytoplasm (Bowles et al., 2005; Sun et al., 2013; Smehilova et al., 2016). To assess the subcellular localization of GmUGT protein, a plasmid carrying *GmUGT* fused with a yellow fluorescent protein (YFP) gene driven by the CaMV 35S promoter was injected into tobacco leaves. The diffused fluorescence of the fusion protein was observed in the cytosol, accumulating along the cell membrane and in the nuclei (Figure 4D), proving GmUGT to be a cytosolic enzyme.

GmUGT Negatively Regulates Flavonoids Content

Because the *Arabidopsis* *ugt72b1* mutant previously displayed an aggravated cell wall lignification and an enhanced flavonoid content (Lin et al., 2016), we investigated whether *GmUGT*, the homolog of ATUGT72B1, plays a similar role in soybeans. First, we evaluated the lignin in the cell walls of the leaves of the *ko-3* and WT plants to find out if they had similar cell wall lignin (Supplementary Figure 4A). To determine the flavonoid content, metabolomics analysis was performed on the *ko-3* and WT plants 36 h after attack or unattack by *H. armigera*; this duration was chosen considering that *GmUGT* gene expression reached its peak after 36 h of attack by *H. armigera* (Figure 4B). Differential metabolite analysis showed that 70 metabolites in unattacked *ko-3* plants mapped to 14 KEGG pathways, while 56 metabolites in the attacked *ko-3* mapped to 10 KEGG pathways (Supplementary Tables 2, 3 and Figures 5A,B). The *GmUGT* encodes a UDP-7-O-glucosyltransferase, which catalyzes the final step of flavonoid biosynthesis to form their glycoside derivatives. Loss of function of *GmUGT* resulted in increased daidzein and formononetin contents, while genistein content was reduced due to the redirection of metabolic flux to other products (Supplementary Figures 4B–D). Among these KEGG pathways, isoflavonoid and flavonoid biosynthesis were significantly enriched in the *ko-3* plants attacked by *H. armigera* after 36 h (Figures 5A,B). Different flavonoid metabolites were upregulated in the *ko-3* mutants 36 h after attack by *H. armigera* (Figure 5C). These were related to insect resistance, including 7,4-dihydroxyflavone, eupatilin, isoformononetin, ononin, and daidzein (Figure 5D). Furthermore, we tested the expression levels of 12 immunity-related genes in *H. armigera* larvae that were fed *ko-3* plant leaves for 7 days to observe that the expression of most of these genes was significantly decreased (Supplementary Figures 4E–O). These results indicate that *GmUGT* mutagenesis affects the flavonoid content in soybean attacked by *H. armigera*.

Differently Expressed Genes in GmUGT Mutants and Wild Type Plants

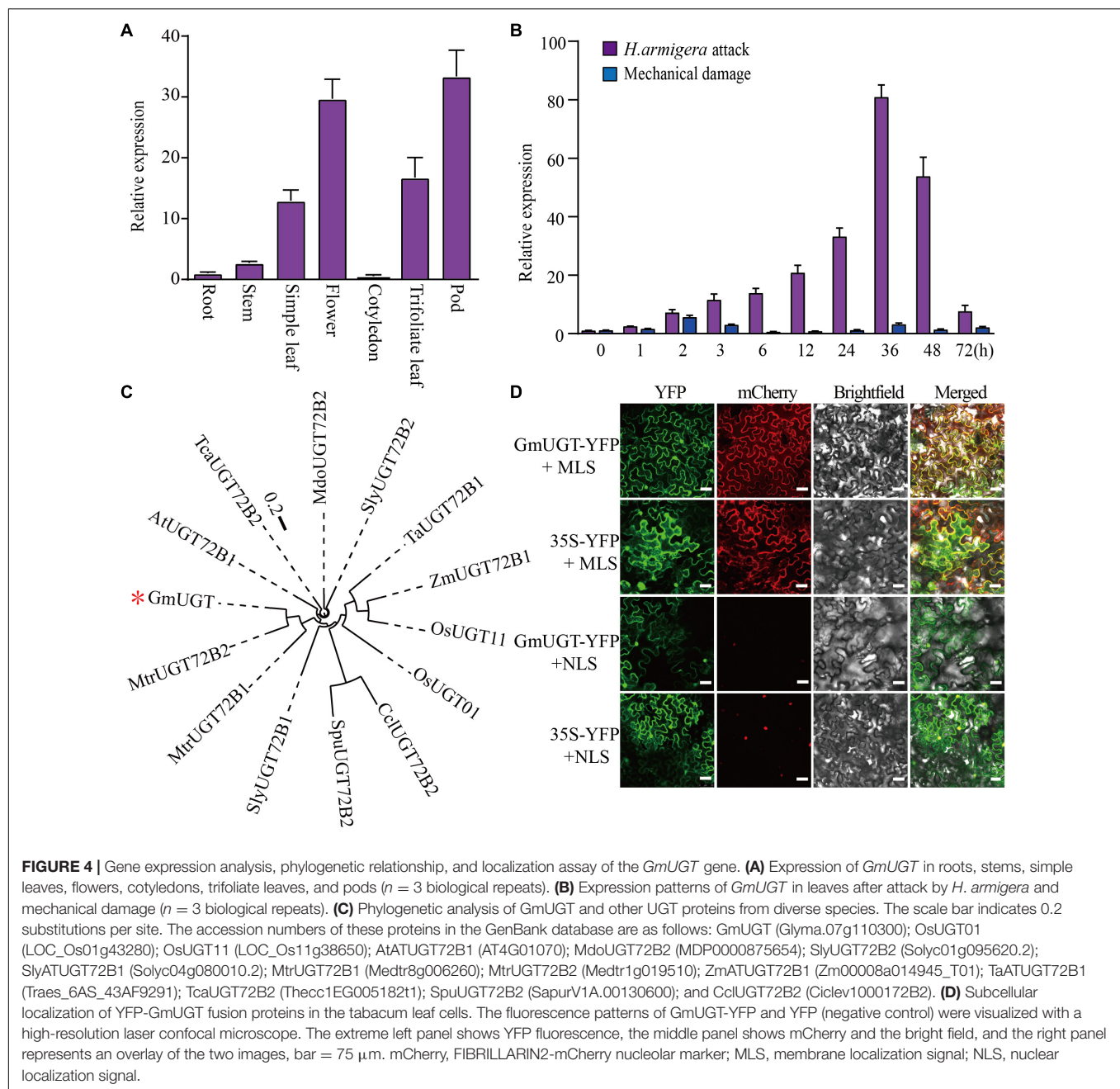
A transcriptome assay was conducted on leaves from the WT, *ko-3* plants that were attacked for 36 h by *H. armigera*, and an unattacked control to identify genes related to resistance against *H. armigera*. The most enriched DEGs in *ko-3* according to the KEGG pathways were associated with isoflavonoid and flavonoid biosynthesis, defensive genes related to plant-pathogen interaction, and mitogen-activated protein kinase (MAPK) signaling pathway (Figures 6A,B, Supplementary Figure 5, and Supplementary Tables 4, 5). The qRT-PCR analysis was employed to investigate the expression of genes involved in flavonoid biosynthesis, including *G. max* CHALCONE SYNTHASE (*GmCHS*), *G. max* Chalcone reductase (*GmCHR*), and *G. max* Cytochrome P450 monooxygenase (*GmCYP81E11*). All three genes were upregulated in *ko-3* as compared to in the WT plants attacked by *H. armigera* (Figures 6C–F). Additionally, genes associated with the defense response were assayed, including *G. max* MITOGEN-ACTIVATED PROTEIN KINASE (*GmMAPK*) and *G. max* PATHOGENESIS RELATED1 (*PR1*). These genes were also upregulated in *ko-3* as compared to in the WT plant attacked by *H. armigera* (Figures 6G–K). The consistency between RNA-Seq and qRT-PCR analysis was evaluated by Pearson's correlation (*R*) analysis based on log2FC revealing an *R* = 0.82 and *p* < 0.0001, which proved the reliability of the RNA-Seq data (Supplementary Figure 5C). These data indicated that the knockout of *GmUGT* enhanced the resistance of soybean against leaf-chewing insects by altering genes related to the flavonoid pathway and defense responses.

Complementation of Resistance to Leaf-Chewing Insects of Arabidopsis ugt72b1 Mutant With GmUGT

Because the GmUGT protein shared high homology with the ATUGT72B1 protein (Figure 4C), we hypothesized that the loss-of-function in the *atugt72b1* mutant would show increased resistance to leaf-chewing insects. The larval feeding assay on the *atugt72b1* mutants and the WT plants was performed using the larvae of *H. armigera*. As expected, the *atugt72b1* mutant showed higher resistance to *H. armigera* than the WT plants (Supplementary Figure 6). To further validate the role of *GmUGT* in imparting resistance to *H. armigera*, we introduced the *GmUGT* gene driven by the ATUGT72B1 promoter into the *atugt72b1* mutant and obtained three transgenic lines (RE-1, RE-2, and RE-3) with elevated expression of *GmUGT* in *Arabidopsis* (Figure 7A). All three transgenic lines showed a substantial reduction in resistance to *H. armigera* (Figure 7), indicating that the UGT gene family is consistently involved in imparting resistance to *H. armigera*.

DISCUSSION

Defoliator insects affect soybean yield and quality, including *H. zea*, *H. armigera*, and *S. litura* (Otuka et al., 2020; Haile et al., 2021). Soybean varieties have been screened for native



resistance to such defoliators (Kim et al., 2013, 2015; Liu et al., 2016; Haile et al., 2021). However, breeding such varieties can be labor-intensive and time-consuming (Ortega et al., 2016). Transgenic breeding is a crucial breakthrough in crops breeding for insect resistance. Nevertheless, only six commercialized soybean cultivars with various insect-resistance genes have been identified (Kumar et al., 2020). Among these soybean cultivars, genetically modified cultivars expressing one or more *Bacillus thuringiensis* (Bt) proteins have been widely deployed for pest management (Kumar et al., 2020). This trait provides effective control management for defoliators, including *H. armigera*. Similar to the transgenic crops randomly integrating foreign

DNA into plant genomes; the durability, environmental safety, and potential adverse health effects of the Bt transgenic soybeans have always been controversial, limiting the transgenic breeding (Gao, 2021). Recently, CRISPR/Cas9-based genome editing has become popular as it allows the generation of crop varieties similar to non-transgenic crops without the addition of any foreign DNA to the genome (Bao et al., 2019). Here, we presented a strategy to increase the resistance against leaf-chewing insects in soybean through CRISPR/Cas9-mediated targeted mutagenesis of the *UGT* genes through the flavonoid biosynthesis pathway, which can help to accelerate the breeding process for insect resistance and minimize the great concern to human health.

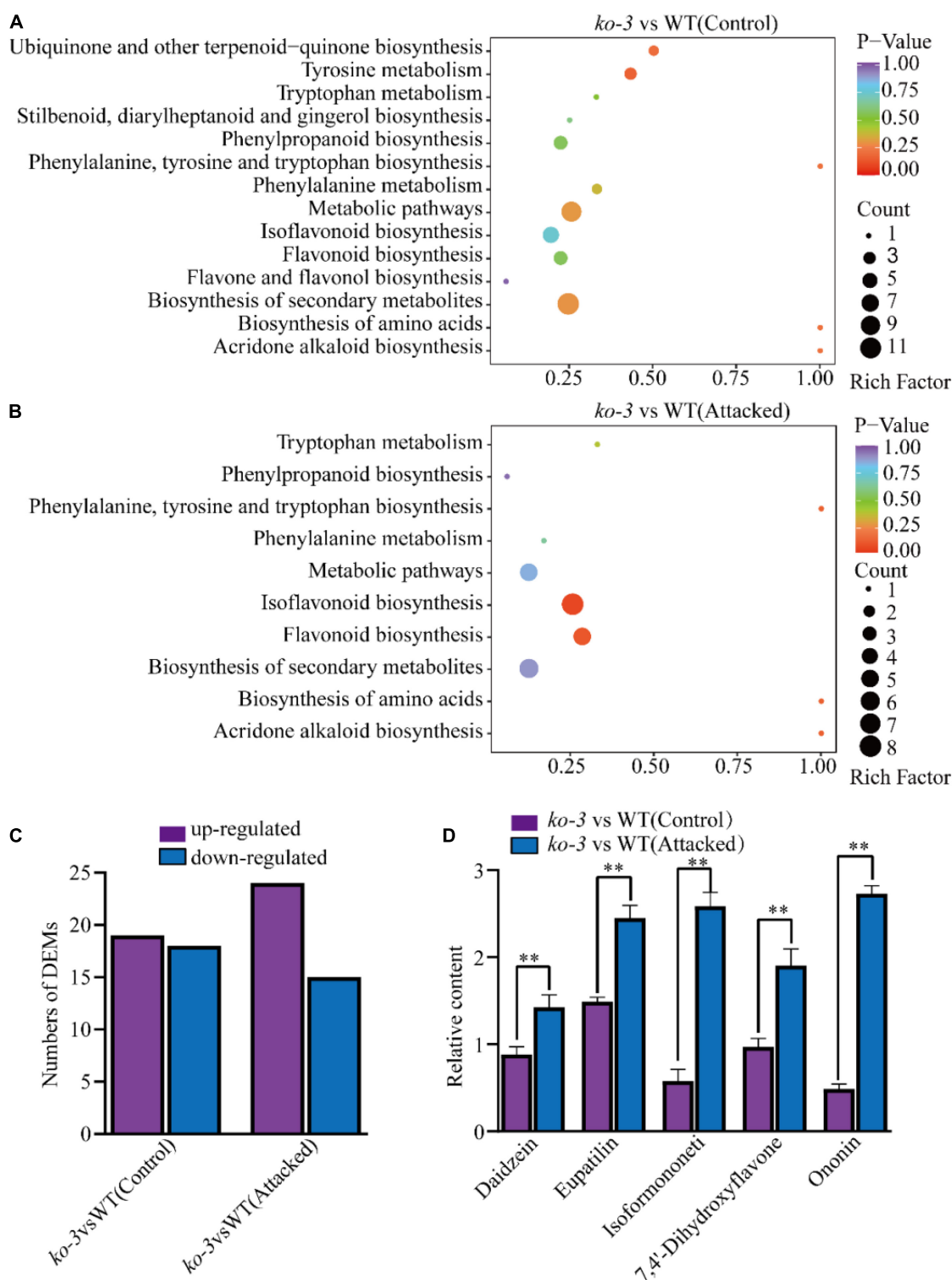


FIGURE 5 | Analysis of different metabolites in *ko-3* versus WT unattacked and attacked by *H. armigera* at 36 h. KEGG pathway enrichment of different metabolites in *ko-3* versus WT plants unattacked by *H. armigera*. **(A)**, and attacked by *H. armigera* at 36 h **(B)**. Each dot represents a KEGG pathway. The Y-axis indicates the KEGG pathway, and the X-axis indicates the rich factor (indicating the degree of differential metabolite enrichment in each pathway; the larger the rich factor is, the greater the differential metabolite enrichment). The dot size and color indicate the number and p value of different metabolites in the pathway, respectively, and a larger dot means more different metabolites. **(C)** The number of different metabolites of flavonoids among *ko-3* versus WT unattacked and attacked by *H. armigera* at 36 h. **(D)** Relative content of metabolites in *ko-3* unattacked and attacked by *H. armigera* at 36 h. The data shown are the means and standard deviations ($n = 3$ biological repeats). Statistically significant differences are marked with asterisks (** $p < 0.01$; Student's *t*-test).

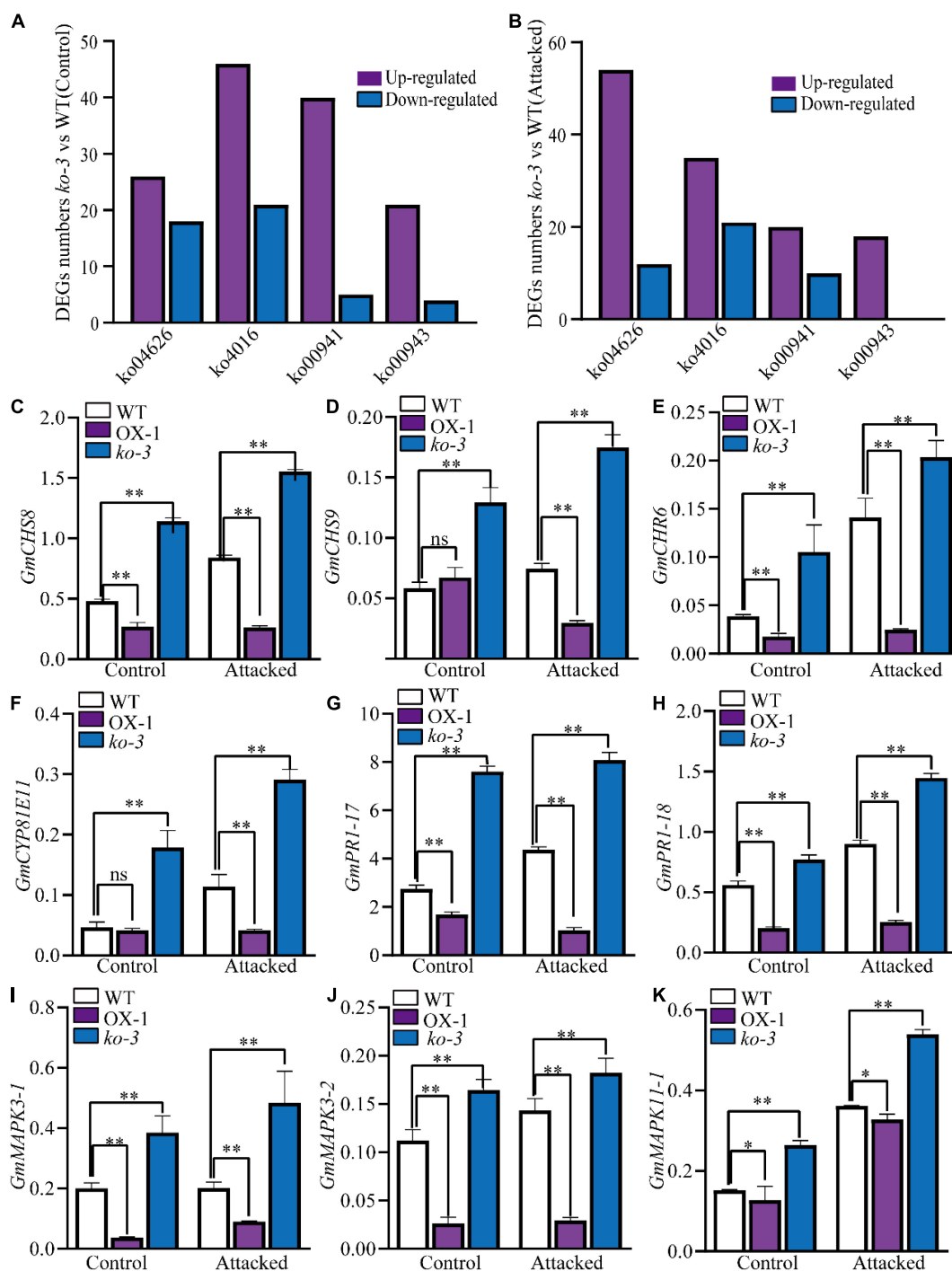


FIGURE 6 | Analysis of differentially expressed genes (DEGs) in *ko-3* versus WT plants unattacked and attacked by *H. armigera* at 36 h. The numbers of DEGs in *ko-3* versus WT plants unattacked by *H. armigera* (A) and attacked by *H. armigera* at 36 h (B). (C–K) Validation of the expression levels of nine DEGs by qRT-PCR. Relative expression levels were normalized to the expression level of *GmSKIP16*. The data shown are the means and standard deviations ($n = 3$ biological repeats). Statistically significant differences are marked with asterisks (ns $p > 0.05$, $0.01 < p < 0.05$, $**p < 0.01$; Student's *t*-test; ns, non-significant).

Ortega et al. (2017) reported that *GmUGT* is a functional SNP marker that exhibits activity against *H. zea*, but information about the *GmUGT* gene and its role in imparting resistance to *H. armigera* remains limited. Here, we first characterized the

GmUGT gene in soybean and observed that it could respond to defoliator attacks rather than mechanical damage (Figure 4B). We observed that the *GmUGT* gene possesses a conserved plant secondary product, the glycosyl-transferase-box (PSPG)

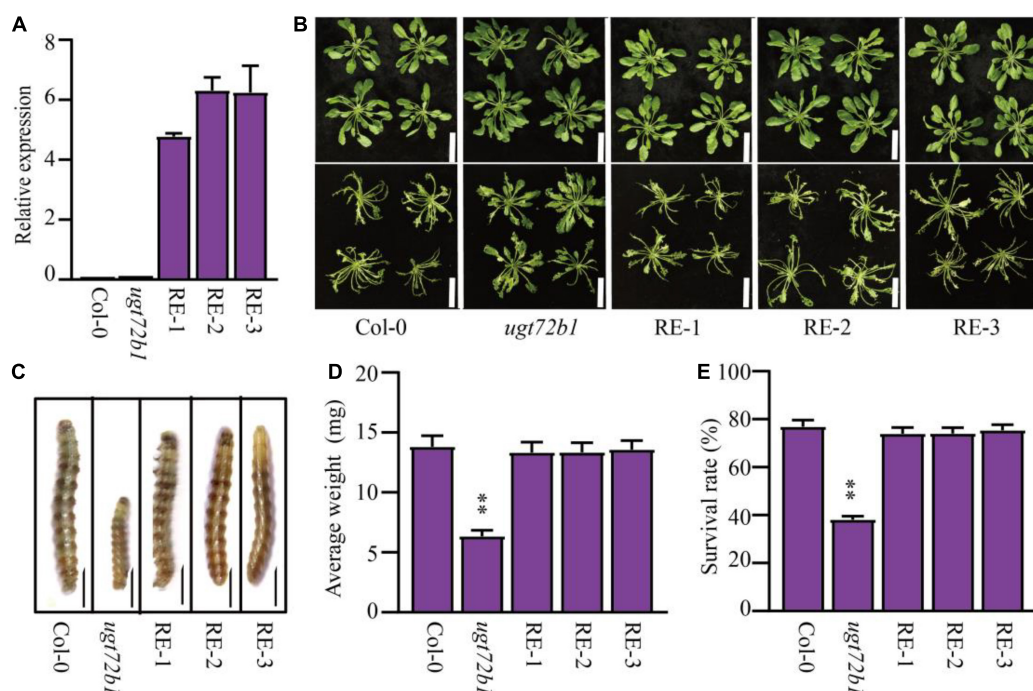


FIGURE 7 | Complementation of resistance to *H. armigera* of *Arabidopsis atugt72b1* mutant with *GmUGT*. **(A)** Confirmation of the expression of *GmUGT* in transgenic *Arabidopsis atugt72b1* mutants. Gene expression levels were normalized with Col-0 (*AtUGT72B1* expression level) as 1. *AtACTIN2* was used as the internal reference control. The data shown are the means and standard deviations ($n = 3$ biological repeats). **(B)** Phenotype of Col-0, *atugt72b1* and three transgenic lines (RE-1, RE-2 and RE-3) attacked by *H. armigera* for 3 days (bar = 2 cm). **(C)** Phenotype of *H. armigera* larvae fed Col-0, *atugt72b1* and three transgenic lines for 7 days (bar = 0.1 cm). **(D)** Average weight of *H. armigera* larvae fed Col-0, *atugt72b1* and three transgenic lines for 7 days. Data shown are the means and standard deviation ($n = 60$ larval). **(E)** The survival rate of *H. armigera* larvae fed Col-0, *atugt72b1* and three transgenic lines for 7 days ($n = 3$ biological replicates, and 100 *H. armigera* larvae in each replicate). Statistically significant differences are marked with asterisks (ns $p > 0.05$, ** $p < 0.01$; Student's *t*-test; ns, non-significant).

motif in its C-terminus (Supplementary Figure 7). The PSPG motif is involved in the recognition and binding of sugar donors (Masada et al., 2007). Additionally, the UGT protein can transfer glycosyl moieties from UDP-sugars to flavonoids during flavonoid biosynthesis (Witte et al., 2009). In soybeans, Yin et al. (2017) identified 212 *UGT* genes and reported that several of them were potentially involved in flavonoid glycosylation. Here, we found that the CRISPR/Cas9-mediated mutagenesis of *GmUGT* led the transgenic plant to exhibit an enhanced resistance to *H. armigera* and *S. litura* (Figure 2). The *Arabidopsis ugt72b1* mutant showed aggravated cell wall lignification and an enhanced flavonoid content (Lin et al., 2016). Both cell wall lignification and flavonoids contribute to resistance against leaf-chewing insects. Cell wall lignification is the first physical barrier against leaf-chewing insects (War et al., 2012). As a secondary metabolite, flavonoids protect plants against leaf-chewing insects by influencing insect behavior, growth, and development (Belete, 2018). Recently, Yang et al. (2021) reported that CRISPR/Cas9-mediated targeted mutagenesis of the *OsUGT707A3* gene in rice (*Oryza sativa*) led to more accumulation of naringenin, lowered naringenin-7-O-b-D-glucoside and apigenin-7-O-b-D-glucoside, and increased the resistance to insects, such as *S. litura*. Thus, we investigated whether the enhanced resistance against leaf-chewing insects due to *GmUGT* mutations was attributable to aggravated lignification of cell walls and alteration

of flavonoids. Our results indicated that the leaf cell wall lignification and *GmUGT* mutations were not significantly related (Supplementary Figure 4A).

Metabolomics analysis revealed that loss of function of *GmUGT* altered flavonoids in soybean leaves (Figure 5). The expression levels of immunity-related genes in *H. armigera* larvae, including pattern recognition receptors (PRRs) and antimicrobial peptides (AMPs) (Wang et al., 2010), decreased significantly in *H. armigera* larvae that were fed with *ko-3* plant leaves for 7 days, indicating that long-term feeding with *ko-3* plant leaves may destroy the immune systems of larvae (Supplementary Figures 4E–O). Therefore, the enhanced resistance of soybean due to *GmUGT* mutations against leaf-chewing insects may be dependent on altered flavonoids.

Transcriptome assay and qRT-PCR analysis showed that the genes involved in flavonoid biosynthesis were upregulated in *ko-3* plants after the *H. armigera* attack. Previous studies have shown that the genes encoding the CHS, CHR, and CYP81E11 enzymes play a critical role in flavonoid biosynthesis (Dhaubhadel et al., 2007; Zhang X. B. et al., 2017; Mamedea et al., 2018; Pi et al., 2019; Zuk et al., 2019). One of the earliest signaling events after leaf-chewing insect attack is the activation of mitogen-activated protein kinases (MAPKs), which are essential to activate the transcription of defense-related genes and accumulate associated metabolites in plants. For example,

MAPKs can upregulate the defense gene, *GmPR-1*, when plants are attacked by insects (Hettenhausen et al., 2015; Giacometti et al., 2016; Breen et al., 2017; Fabricio and Venancio, 2021). Additionally, flavonoids that function as signal molecules may directly interfere with MAPK activities and profoundly affect signaling systems (Agati et al., 2013). Consistent with this, we found out that several genes that were involved in the MAPK pathway, along with the *GmPR-1* genes, were upregulated in *ko-3* compared with the WT plants upon attack by *H. armigera* (Figure 6). These results indicate that the *GmUGT* gene plays an essential role in controlling resistance to leaf-chewing insects through the flavonoid biosynthesis pathway in soybeans.

Interestingly, the *Arabidopsis ugt72b1* mutant also exhibited increased resistance to *H. armigera* (Supplementary Figure 6), while GmUGT shared high homology with AtUGT72B1 (Figure 4C). Additionally, our results showed that the ectopic expression of the *GmUGT* gene in the *Arabidopsis ugt72b1* mutant substantially lessened the resistance of *Arabidopsis ugt72b1* mutant to *H. armigera* (Figure 7). Lin et al. (2016) reported that *ugt72b1* mutant exhibited growth repression after bolting, while the *ugt72b1* mutant phenotype was restored to the WT by expression of AtUGT72B1. However, we observed that the *Arabidopsis ugt72b1* mutant was not restored to its WT by expressing *GmUGT* (Supplementary Figure 8). Furthermore, mutations and overexpression of *GmUGT* did not cause any noticeable phenotypic changes (Supplementary Figures 1C, 3A). These results indicate that the two homologous genes, AtUGT72B1 and *GmUGT*, are functionally differentiated, but they play a conserved role in imparting resistance to leaf-chewing insects.

CONCLUSION

The CRISPR/Cas9-mediated mutagenesis of *GmUGT* increased soybean resistance to *H. armigera* and *S. litura* by altering flavonoid content and by changing the expression of genes related to flavonoid biosynthesis and defense response. The *atugt72b1* mutant of *Arabidopsis* also exhibited an increased resistance to *H. armiger*, while the ectopic expression of the *GmUGT* gene in this mutant substantially reduced the resistance against

H. armigera. Our findings highlight an effective strategy for increasing resistance to leaf-chewing insects in soybean through CRISPR/Cas9-mediated targeted mutagenesis of the *UGT* genes, which may also work for other crops.

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 below: <https://www.ncbi.nlm.nih.gov/>, SAMN21448187, SAMN21448188, SAMN21448189, SAMN21448190, SAMN21448191, SAMN21448192, SAMN21448193, SAMN21448194, SAMN21448195, SAMN21448196, and SAMN21448197.

AUTHOR CONTRIBUTIONS

YZ, WG, DC, XL, and YJ designed the experiments. YZ, WG, LC, XS, HY, YF, WO, SM, HC, SC, QH, YH, DQ, ZS, ZY, SY, CZ, and XZ performed the experiments and analyzed the data. YZ, WG, DC, and XL wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The research was supported by The National Key Research and Development Program of China (2016YFD0100603); The National Natural Science Foundation of China (32072087); and The Science and Technology Innovation Project of Chinese Academy of Agricultural Sciences, Cultivation of New Science and Technology (2021-2060302-049-022).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.802716/full#supplementary-material>

REFERENCES

- Agati, G., Brunetti, C., Ferdinando, M. D., Ferrini, F., Pollastri, S., and Tattini, T. (2013). Functional roles of flavonoids in photoprotection: new evidence, lessons from the past. *Plant. Physiol. Bioch.* 72, 35–45. doi: 10.1016/j.plaphy.2013.03.014
- Aljibory, Z., and Chen, M. S. (2018). Indirect plant defense against insect herbivores: a review. *Insect. Sci.* 25, 2–23. doi: 10.1111/1744-7917.12436
- Arifin, H. A., Hashiguchi, T., Nagahama, K., Hashiguchi, M., Muguerza, M., Sakakibara, Y., et al. (2021). Varietal differences in flavonoid and antioxidant activity in Japanese soybean accessions. *Biosci. Biotechnol. Biochem.* 85, 916–922. doi: 10.1093/bbb/zbab104
- Bai, G., Yang, D. H., Zhao, Y., Ha, S., Yang, F., Ma, J., et al. (2013). Interactions between soybean ABA receptors and type 2C protein phosphatases. *Plant. Mol. Biol.* 83, 651–664. doi: 10.1007/s11103-013-0114-4
- Bao, A., Chen, H. F., Chen, L. M., Chen, S. L., Hao, Q. N., Guo, W., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. *BMC Plant. Biol.* 19:131. doi: 10.1186/s12870-019-1746-6
- Belete, T. (2018). Defense mechanisms of plants to insect pests: from morphological to biochemical approach. *Trends. Tech. Sci. Res.* 2, 30–38. doi: 10.1080/15548627.2020.1797280
- Bondori, A., Bagheri, A., Damalas, C. A., and Allahyari, M. S. (2018). Use of personal protective equipment towards pesticide exposure: farmers' attitudes and determinants of behavior. *Sci. Total. Environ.* 639, 1156–1163. doi: 10.1016/j.scitotenv.2018.05.203
- Bowles, D., Isayenkova, J., Lim, E. K., and Poppenberger, B. (2005). Glycosyltransferases: managers of small molecules. *Curr. Opin. Plant. Biol.* 8, 254–263. doi: 10.1016/j.pbi.2005.03.007
- Breen, S., Williams, S. J., Outram, M., Kobe, B., and Solomon, P. S. (2017). Emerging insights into the functions of pathogenesis-related protein 1. *Trends. Plant. Sci.* 22, 871–879. doi: 10.1016/j.tplants.2017.06.013
- Cai, Y. P., Wang, L. W., Chen, L., Wu, T. T., Liu, L. P., Sun, S., et al. (2020). Mutagenesis of *GmFT2a* and *GmFT5a* mediated by CRISPR/Cas9 contributes

- for expanding the regional adaptability of soybean. *Plant. Biotechnol. J.* 18, 298–309. doi: 10.1111/pbi.13199
- Chen, K., Wang, Y. W., Zhang, R., Zhang, H., and Gao, C. (2019). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant. Biol.* 70, 667–697.
- Chen, L., Yang, H., Fang, Y., Guo, W., Chen, H., Zhang, X., et al. (2021). Overexpression of *GmMYB14* improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnol. J.* 19, 702–716. doi: 10.1111/pbi.13496
- Chen, W., Gong, L., Guo, Z. L., Wang, S. H., Zhang, H. Y., Liu, X. Q., et al. (2013). A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: application in the study of rice metabolomics. *Mol. Plant.* 6, 1769–1780. doi: 10.1093/mp/ss.t080
- Chen, Y. K., Xu, X. P., Liu, Z. X., Zhang, Z. H., Han, X., Lin, Y. L., et al. (2020). Global scale transcriptome analysis reveals differently expressed genes involve in early somatic embryogenesis in *Dimocarpus longan* Lour. *BMC Genomics*. 21:4. doi: 10.1186/s12864-019-6393-7
- Cheng, Q., Dong, L. D., Su, T., Li, T. Y., Gan, Z. R., Nan, H. Y., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of *GmLHY* genes alters plant height and internode length in soybean. *BMC Plant. Biol.* 19:562. doi: 10.1186/s12870-019-2145-8
- Clough, S. J., and Bent, A. F. (1998). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant. J.* 16, 735–743.
- Deguchi, M., Bogush, D., Weeden, H., Spuhler, Z., Potlakayala, S., Kondo, T., et al. (2020). Establishment and optimization of a hemp (*Cannabis sativa* L.) agroinfiltration system for gene expression and silencing studies. *Sci. Rep.* 10:3504. doi: 10.1038/s41598-020-60323-9
- Dhaubhadel, S., Gijzen, M., Moy, P., and Farhangkhoei, M. (2007). Transcriptome analysis reveals a critical role of *CHS7* and *CHS8* genes for isoflavonoid synthesis in soybean seeds. *Plant. Physiol.* 143, 326–338. doi: 10.1104/pp.106.086306
- Domingo, J. L., and Bordonaba, J. G. (2011). A literature review on the safety assessment of genetically modified plants. *Environ. Int.* 37, 734–742. doi: 10.1016/j.envint.2011.01.003
- Du, H. Y., Zeng, X. R., Zhao, M., Cui, X. P., Wang, Q., Yang, H., et al. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J. Biotechnol.* 217, 90–97. doi: 10.1016/j.jbiotec.2015.11.005
- Elba, D. L. C., Viria, B. D., Ramirez, F., and Castillo, L. E. (2014). Environmental hazards associated with pesticide import into Costa Rica, 1977–2009. *J. Environ. Biol.* 35, 43–55.
- Fabricio, A. S., and Venancio, T. M. (2021). Pathogenesis-related protein 1 (PR-1) genes in soybean: genome-wide identification, structural analysis and expression profiling under multiple biotic and abiotic stresses. *Biorxiv [preprint]* 5:23. doi: 10.1101/2021.03.27.437342
- Gao, C. X. (2021). Genome engineering for crop improvement and future agriculture. *Cell.* 184, 1621–1635. doi: 10.1016/j.cell.2021.01.005
- Ghione, C. E., Lombardo, L. A., Vicentin, I. G., and Heinz, R. A. (2021). Association mapping to identify molecular markers associated with resistance genes to stink bugs in soybean. *Euphytica* 217, 46–58.
- Giacometti, R., Barneto, J., Barriga, L. G., Sardoy, P. M., Balestrasse, K., Andrade, A. M., et al. (2016). Early perception of stink bug damage in developing seeds of field-grown soybean induces chemical defences and reduces bug attack. *Pest. Manag. Sci.* 72, 1585–1594. doi: 10.1002/ps.4192
- Haile, F., Nowatzki, T., Storer, N., and Davis, J. (2021). Overview of pest status, potential risk, and management considerations of *Helicoverpa armigera* (Lepidoptera: noctuidae) for U.S. soybean production. *J. Integr. Pest. Manag.* 12, 1–10.
- Han, J. N., Guo, B. F., Guo, Y., Zhang, B., Wang, X. B., and Qiu, L. J. (2019). Creation of early flowering germplasm of soybean by CRISPR/Cas9 technology. *Front. Plant. Sci.* 10:1446. doi: 10.3389/fpls.2019.01446
- Hettenhausen, C., Schuman, M. C., and Wu, J. Q. (2015). MAPK signaling: a key element in plant defense response to insects. *Insect. Sci.* 22, 157–164. doi: 10.1111/1744-7917.12128
- Ichim, M. C. (2019). The Romanian experience and perspective on the commercial cultivation of genetically modified crops in Europe. *Transgenic. Res.* 28, 1–7. doi: 10.1007/s11248-018-0095-9
- Kim, H., Xing, G. G., He, J. B., Zhao, T. J., Yang, S. P., Li, Y., et al. (2015). An environmental different association analysis of antibiosis to common cutworm in a Chinese soybean germplasm population and optimization of the cross design. *Mol. Breeding*. 35:76. doi: 10.1007/s11032-015-0267-8
- Kim, H., Xing, G. G., Wang, Y. F., Zhao, T. J., Yu, D. Y., Yang, S. P., et al. (2013). Constitution of resistance to common cutworm in terms of antibiosis and antixenosis in soybean RIL populations. *Euphytica* 196, 137–154. doi: 10.1007/s10681-013-1021-0
- Knott, G. J., and Jennifer, A. D. (2018). CRISPR-Cas guides the future of genetic engineering. *Science*. 361, 866–869. doi: 10.1126/science.aat5011
- Kou, M. Z., Bastias, D. A., Christensen, M. J., Zhong, R., Nan, Z. B., and Zhang, X. X. (2021). The plant salicylic acid signalling pathway regulates the infection of a biotrophic pathogen in grasses associated with an *Epichloe* endophyte. *J. Fungi*. 7, 633–650. doi: 10.3390/jof7080633
- Kumar, K., Gambhir, G., Dass, A., Tripathi, A. K., Singh, A., Jha, A. K., et al. (2020). Genetically modified crops: current status and future prospects. *Planta*. 251:91. doi: 10.1007/s00425-020-03372-8
- Lei, X. J., Tan, B., Liu, Z. Y., Wu, J., Lv, J. X., and Gao, C. Q. (2021). *ThCOL2* improves the salt stress tolerance of *Tamarix hispida*. *Front. Plant. Sci.* 12:653791. doi: 10.3389/fpls.2021.653791
- Lin, J. S., Huang, X. X., Li, Q., Cao, Y., Bao, Y., Meng, X. F., et al. (2016). UDP-glycosyltransferase 72B1 catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. *Plant. J.* 88, 26–42. doi: 10.1111/tpj.13229
- Liu, H. J., Che, Z. J., Zeng, X. R., Zhang, G. Z., Wang, H., and Yu, D. Y. (2016). Identification of single nucleotide polymorphisms in soybean associated with resistance to common cutworm (*Spodoptera litura* Fabricius). *Euphytica*. 209, 49–62.
- 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
- Lucini, T. A., Panizzi, A. R., and Bueno, A. D. F. (2021). Evaluating resistance of the soybean block technology cultivars to the Neotropical brown stink bug, *Euschistus heros* (F.). *J. Insect. Physiol.* 131:104228. doi: 10.1016/j.jinsphys.2021.104228
- Mamede, R. T., Waki, Y., Kawai, S., Takahashi, S., and Nakayama, T. (2018). Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of *GmCHRS5*, which interacts with 2-hydroxyisoflavanone synthase. *Plant. J.* 96, 56–74. doi: 10.1111/tpj.14014
- Masada, S., Terasaka, K., and Mizukami, H. (2007). A single amino acid in the PSPG-box plays an important role in the catalytic function of CaUGT2 (Curcumin glucosyltransferase), a Group D Family 1 glucosyltransferase from *Catharanthus roseus*. *FEBS. Letters*. 581, 2605–2610. doi: 10.1016/j.febslet.2007.05.002
- Missbach, S., Weis, B. L., Martin, R., Simm, S., Bohnsack, M. T., and Schleiff, E. (2013). 40S ribosome biogenesis co-factors are essential for gametophyte and embryo development. *PLoS One* 8:e54084. doi: 10.1371/journal.pone.0054084
- Ortega, M. A., All, J. N., Boerma, H. R., and Parrott, W. A. (2016). Pyramids of QTLs enhance host-plant resistance and Bt-mediated resistance to leaf-chewing insects in soybean. *Theor. Appl. Genet.* 129, 703–715. doi: 10.1007/s00122-015-2658-y
- Ortega, M. A., Lail, L. A., Wood, E. D., All, J. N., Li, Z., Boerma, H. R., et al. (2017). Registration of two soybean germplasm lines containing leaf-chewing insect resistance QTLs from PI 229358 and PI 227687 introgressed into 'Benning'. *J. Plant. Regist.* 11, 185–191. doi: 10.3198/jpr2016.04.0019crg
- Otuka, A. M., Matsumura, M., and Tokuda, M. (2020). Dispersal of the common cutworm, *spodoptera litura*, monitored by searchlight trap and relationship with occurrence of soybean leaf damage. *Insects* 11:427. doi: 10.3390/insects11070427
- Pi, E., Xu, J., Li, H. H., Fan, W., Zhu, C. M., Zhang, T. Y., et al. (2019). Enhanced salt tolerance of rhizobia-inoculated soybean correlates with decreased phosphorylation of the transcription factor *GmMYB183* and altered flavonoid biosynthesis. *Mol. Cell. Proteomics* 18, 2225–2243. doi: 10.1074/mcp.RA119.001704
- Pozebon, H., Marques, R. P., Padilha, G., Neal, M. O., Valmorbidia, I., Bevilacqua, J. G., et al. (2020). Arthropod invasions versus soybean production in Brazil: a review. *J. Econ. Entomol.* 113, 1591–1608. doi: 10.1093/jeet/toaa108

- Raman, R. (2017). The impact of genetically modified (GM) crops in modern agriculture: a review. *GM Crops. Food* 8, 195–208. doi: 10.1080/21645698.2017.1413522
- Rector, B., All, J., Parrott, W., and Boerma, H. R. (2000). Quantitative trait loci for antibiosis resistance to corn earworm in soybean. *Crop. Sci.* 40, 233–238. doi: 10.2135/cropsci2000.401233x
- Sabljić, I., Barneto, J. A., Balestrasse, K. B., Zavala, J. A., and Pagano, E. A. (2020). Role of reactive oxygen species and isoflavonoids in soybean resistance to the attack of the southern green stink bug. *Peer. J.* 8:e9956. doi: 10.7717/peerj.9956
- Saikhedkar, N. S., Joshi, R. S., Yadav, A. K., Seal, S., Fernandes, M., and Giri, A. P. (2019). Phyto-inspired cyclic peptides derived from plant Pin-II type protease inhibitor reactive center loops for crop protection from insect pests. *Biochim. Biophys. Acta. Gen. Subj.* 1863, 1254–1262. doi: 10.1016/j.bbagen.2019.05.003
- Schiemann, J., Dietz-Pfeilstetter, A., Hartung, F., Kohl, C., Romeis, J., and Sprink, T. (2019). Risk assessment and regulation of plants modified by modern biotechniques: current status and future challenges. *Annu. Rev. Plant. Biol.* 70, 699–726. doi: 10.1146/annurev-arplant-050718-100025
- Shen, X. J., Wang, Y. Y., Zhang, Y. X., Guo, W., Jiao, Y. Q., and Zhou, X. A. (2018). Overexpression of the wild soybean R2R3-MYB transcription factor GsMYB15 enhances resistance to salt stress and *Helicoverpa armigera* in transgenic *Arabidopsis*. *Int. J. Mol. Sci.* 19:3958. doi: 10.3390/ijms19123958
- Smehilova, M., Dobruskova, J., Novak, Q., Takac, T., and Galuszka, P. (2016). Cytokinin-specific glycosyltransferases possess different roles in cytokinin homeostasis maintenance. *Front. Plant. Sci.* 7:1264. doi: 10.3389/fpls.2016.01264
- Sun, Y. G., Wang, B., Jin, S. H., Qu, X. X., Li, Y. J., and Hou, B. K. (2013). Ectopic expression of *Arabidopsis* glycosyltransferase UGT85A5 enhances salt stress tolerance in tobacco. *PLoS One* 8:e59924. doi: 10.1371/journal.pone.0059924
- Thevenot, E. A., Roux, A., Xu, Y., Ezan, E., and Junot, C. (2015). Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J. Proteome. Res.* 14, 3322–3335. doi: 10.1021/acs.jproteome.5b00354
- Wang, H., Yan, H. L., Du, H. P., Chao, M. N., Gao, Z. J., and Yu, D. Y. (2015). Mapping quantitative trait loci associated with soybean resistance to common cutworm and soybean compensatory growth after defoliation using SNP marker-based genome-wide association analysis. *Mol. Breeding*. 35:168. doi: 10.1016/j.cj.2014.08.004
- Wang, Q., Liu, Y., He, H. J., Zhao, X. F., and Wang, J. X. (2010). Immune responses of *Helicoverpa armigera* to different kinds of pathogens. *BMC Immunol.* 11:9. doi: 10.1186/1471-2172-11-9
- War, A. R., Paulraj, M. G., Ahmad, T., Buhroo, A. A., Hussain, B., Ignacimuthu, S., et al. (2012). Mechanisms of plant defense against insect herbivores. *Plant. Signal. Behav.* 7, 1306–1320. doi: 10.4161/psb.21663
- Witte, S., Moco, S., Vervoort, J., Matern, U., and Martens, S. (2009). Recombinant expression and functional characterisation of regiospecific flavonoid glycosyltransferases from *Hieracium pilosella* L. *Planta* 229, 1135–1146. doi: 10.1007/s00425-009-0902-x
- Yang, Z. Y., Li, N., Kitano, T., Li, P., Spindel, J. E., Wang, L. S., et al. (2021). Genetic mapping identifies a rice naringenin O-glucosyltransferase that influences insect resistance. *Plant. J.* 106, 1401–1413. doi: 10.1111/tip.15244
- Yin, Q. G., Shen, G. A., Di, S. K., Fan, C. Y., Chang, Z. Z., and Pang, Y. Z. (2017). Genome-wide identification and functional characterization of UDP-glucosyltransferase genes involved in flavonoid biosynthesis in *Glycine max*. *Plant. Cell. Physiol.* 58, 1558–1572. doi: 10.1093/pcp/pcx081
- Zhang, X. B., Abrahán, C., Colquhoun, T. A., and Liu, C. J. (2017). A proteolytic regulator controlling chalcone synthase stability and flavonoid biosynthesis in *Arabidopsis*. *Plant. Cell.* 29, 1157–1174. doi: 10.1105/tpc.16.00855
- Zhang, X. L., Zhou, Q. X., Zou, W., and Hu, X. G. (2017). Molecular mechanisms of developmental toxicity induced by graphene oxide at predicted environmental concentrations. *Environ. Sci. Technol.* 51, 7861–7871. doi: 10.1021/acs.est.7b01922
- Zhu, S., Saski, C. A., Boerma, H. R., Tomkins, J. P., All, J. N., and Parrott, W. A. (2008). Construction of a BAC library for a defoliating insect-resistant soybean and identification of candidate clones using a novel approach. *Plant Mol. Biol. Rep.* 27, 229–235. doi: 10.1007/s11105-008-0077-9
- Zhu, S., Walker, D. R., Boerma, H. R., All, J. N., and Parrott, W. A. (2006). Fine mapping of a major insect resistance QTL in soybean and its interaction with minor resistance QTLs. *Crop. Sci.* 46, 1094–1099.
- Zuk, M., Szperlik, J., Hnitecka, A., and Szopa, J. (2019). Temporal biosynthesis of flavone constituents in flax growth stages. *Plant. Physiol. Biochem.* 142, 234–245. doi: 10.1016/j.plaphy.2019.07.009

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, Guo, Chen, Shen, Yang, Fang, Ouyang, Mai, Chen, Chen, Hao, Yuan, Zhang, Huang, Shan, Yang, Qiu, Zhou, Cao, Li and Jiao. 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.



Traits Expansion and Storage of Soybean Phenotypic Data in Computer Vision-Based Test

Yongchao Xing¹, Peixin Lv¹, Hong He^{1*}, Jiantian Leng², Hui Yu² and Xianzhong Feng²

¹ School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China, ² Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China

OPEN ACCESS

Edited by:

Milena Petriccione,
Council for Agricultural Research
and Economics, Italy

Reviewed by:

Yumou Qiu,
Iowa State University, United States
Macon Nardino,
Universidade Federal de Viçosa, Brazil

*Correspondence:

Hong He
hehong@sdu.edu.cn

Specialty section:

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

Received: 10 December 2021

Accepted: 02 February 2022

Published: 01 March 2022

Citation:

Xing Y, Lv P, He H, Leng J, Yu H
and Feng X (2022) Traits Expansion
and Storage of Soybean Phenotypic
Data in Computer Vision-Based Test.
Front. Plant Sci. 13:832592.
doi: 10.3389/fpls.2022.832592

Phenotypic traits of crops are an important basis for cultivating new crop varieties. Breeding experts expect to use artificial intelligence (AI) technology and obtain many accurate phenotypic data at a lower cost for the design of breeding programs. Computer vision (CV) has a higher resolution than human vision and has the potential to achieve large-scale, low-cost, and accurate analysis and identification of crop phenotypes. The existing criteria for investigating phenotypic traits are oriented to artificial species examination, among these are a few traits type that cannot meet the needs of machine learning even if the data are complete. Therefore, the research starts from the need to collect phenotypic data based on CV technology to expand, respectively, the four types of traits in the “Guide to Plant Variety Specificity, Consistency and Stability Testing: Soybean”: main agronomic traits in field investigation, main agronomic traits in the indoor survey, resistance traits, and soybean seed phenotypic traits. This paper expounds on the role of the newly added phenotypic traits and shows the necessity of adding them with some instances. The expanded traits are important additions and improvements to the existing criteria. Databases containing expanded traits are important sources of data for Soybean AI Breeding Platforms. They are necessary to provide convenience for deep learning and support the experts to design accurate breeding programs.

Keywords: soybean test, phenotypic data, computer vision, traits expansion and storage, soybean AI breeding platform

INTRODUCTION

In recent years, artificial intelligence (AI) technology has been widely used in the field of agronomy, which is used to image classification, object recognition, and feature extraction (Pound et al., 2018). The difficulty of plant science research is gradually changing from gene analysis to phenotypic analysis (Mir et al., 2019). The correlation analysis between Phenomics and other omics can analyze the biological laws of crops and effectively serve agricultural production (Zhou et al., 2018). Hu et al. (2019) explains the development of plant phenotypes, and explains in detail the importance of phenotypic collection techniques and image data analysis methods to promote plant phenotype research; Ma (2020) uses computer vision (CV) technology to extract the phenotypic traits of soybean seeds; Ren et al. (2020) proposes a model named Deconvolution-GuidedVGGNet to realize plant leaf disease species identification and spot segmentation at the same time; Guo et al. (2021) use the improved yolov4 to detect the pods of soybean plants.

However, soybean has relatively few phenotypic traits. In the preliminary study of phenotypic traits of soybean varieties using CV, only the plant height and the phenotypic traits of leaf parts are extracted. The preliminary study could not provide enough data to meet the needs of AI breeding. In this study, phenotypic traits of soybean at various stages are expanded based on “Guide to Plant Variety Specificity, Consistency and Stability Testing: Soybean” (Ministry of Agriculture and Rural Affairs of the People’s Republic of China, 2018), using CV processing methods and AI technology, a sufficiently large amount of data is obtained to establish a quantitative analysis technology for the main agronomic traits of different soybean varieties in the field investigation, indoor test, disease resistance, and quality.

QUANTITATIVE ANALYSIS OF SOYBEAN FIELD SURVEY TRAITS BASED ON COMPUTER VISION TECHNOLOGY

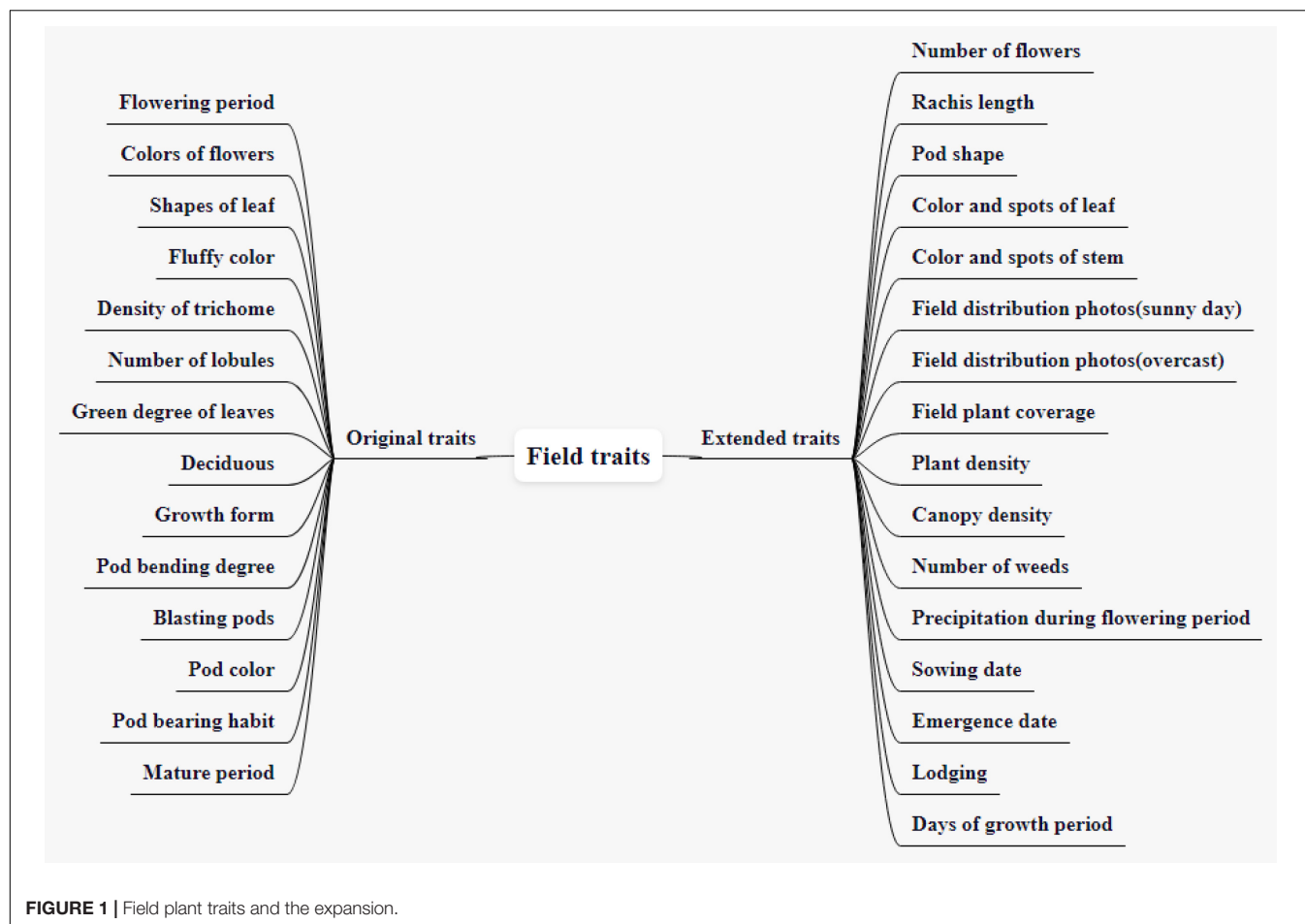
The CV-based phenotypic data acquisition methods could find and remember more details than that were seen to the naked eye, so the currently common 14 field traits are expanded to 21 different traits, as shown in **Figure 1**.

Common Field Investigation Traits

There are 14 original field traits, including flowering period, colors of flowers, shapes of leaf, fluffy color, the density of trichome, number of lobules, green degree of leaves, deciduous, growth form, pod bending degree, blasting pods, pod color, pod bearing habit, and mature period. The sowing date refers to the date on the day of sowing, expressed as month-day. The flowering period refers to the date when 50% of the plants begin to bloom, expressed as month-day. The colors of flowers refer to the color of petals, divided into white and purple. Shapes of the leaf refer to the shape of the middle leaf of the three mature compound leaves in the middle and upper part of the investigated plant during the peak flowering period. It is divided into four types: needle-shaped, triangular, apical ovate, and round ovate (In the north, it is generally pointed leaves, while in the south, it is generally circular leaves. In the north, the sunshine is short, and the light is relatively insufficient, thus, avoid shielding). The fluffy color refers to the color of the fluff on the upper part of the stem or the pod skin of the investigated plant when it is mature, divided into gray and brown (The fluffy color is generally gray in the northeast and brown in the south). The density of fine hair is divided into nine levels, which are extremely thin, extremely thin to thin, thin, thin to medium, medium, medium to dense, dense, dense to extremely dense, and extremely dense. The number of lobules is divided into three lobules, five lobules, and many lobules. The green degree of leaves can be divided into nine types: very light, very light to light, light, light to medium, medium, medium to deep, deep, deep to very deep, very deep. There are three kinds of deciduous traits: non-deciduous, semi-deciduous, and deciduous. According to the growth form of the main stem, it can be divided into semi-straggling, straggling, semi-erect, and erect, which can be used to establish the prediction model of yield per unit area. Pod bending degrees can be divided into four types:

none or very weak, weak, medium, and strong. There are 9 kinds of blasting pods: none or very light, none or very light to light, light, light to medium, medium, medium to heavy, heavy, heavy to very heavy, and very heavy. Pod color is the ripening pod color, divided into white yellow, light yellow, light brown, medium brown, dark brown, and black, which is a total of six colors. Pod-bearing habits are divided into infinite, finite, and sub-finite. Among them, the infinite pod-bearing habit means that new leaves can be produced at the top of the main stem when it blooms. The flowering and pod-setting sequence is from bottom to top, with small top leaves, short inflorescences, scattered pods, and generally 1–2 pods at the top of the main stem. Finite pod-bearing habit means that no new leaves appear on the main stem when it blooms, and there are obvious inflorescences at the top and pod-bearing clusters. The sub-finite pod-setting habit means that the growth traits and pod-setting status at the top of the main stem are between infinite and finite. Generally, 3–4 pods grow on the top of the main stem. The mature period refers to the date when 95% of the pods of the whole plant become mature color, and the plants that start to make a noise when shaking reach more than 50%, expressed in month-day.

The device for acquiring soybean field images is DJI Phantom 4Pro, which has strong battery life, a 1-inch, 20-megapixel sensor, and a dynamic range close to 12 stops. It has better image quality in terms of the level of detail and low-light conditions. Adapt to the changing light intensity in the field. The colors of flowers and shapes of leaves in the original survey standard can only be observed by naked eyes, and the flower color can only be divided into white and purple. To record the flower color more accurately and specifically, it is necessary to use CV technology to process the collected images to obtain the average RGB value of the flower. The RGB is the most commonly used color space, consisting of three-color channels: red (R), green (G), and blue (B). Soybean images in the field under natural conditions are easily affected by light and obstructions and are very sensitive to brightness. In the RGB color space, these three components are highly correlated. The change of brightness will cause the three components to change accordingly, and the naked eye has different perceptions of the three colors. In the HSV color space, H refers to hue, S refers to saturation, and V refers to a degree. It is closer to people’s experience of color perception than RGB, more intuitively expresses the hue, vividness, and brightness of colors, and is more convenient for color contrast. Therefore, in image processing, first, convert the color space of field soybean image from RGB color space to HSV color space, and then, define the range of colors of flowers in HSV color space, generate a mask according to the threshold, and perform bitwise and operation between the mask and the original image. The flower part is extracted from the image, and finally, the color space is converted to RGB, and the RGB average value of the non-zero pixels is extracted as the colors of flowers. The shapes of the leaves are obtained after the edge detection using the Laplacian of the Gaussian operator. The Laplacian of Gaussian is composed of Gaussian filtering and Laplacian. It is an edge detection operator based on the second-order differential method. The Laplacian operator can highlight the areas where the intensity changes rapidly in the image, so, it is often used in edge detection. Since

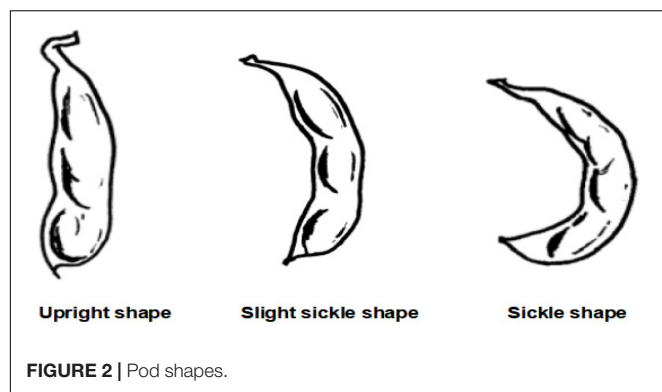


the Laplacian operator is more sensitive to discrete points and noise, it is necessary to use a Gaussian smoothing filter to smooth the image before performing the Laplacian operation to reduce the sensitivity of the Laplacian operation to noise and improve the robustness to noise and discrete points. The following are the Algorithm steps: ①Use Gaussian smoothing filter for smoothing; ②Use Laplacian to calculate the second derivative; ③Detect the zero-crossing point in the image; and ④Set the zero-crossing point as the threshold, and only keep those strong crossing points (there is a big gap between the positive maximum and the negative minimum).

Field Investigation Traits Extended by Computer Vision Techniques

The extended traits of field traits include sowing date, emergence date, lodging, days of the growth period, number of flowers, rachis length, pod shape, color and spots of leaf, color, and spots of the stem, field distribution photographs (sunny day), field distribution photographs (overcast), field plant coverage, plant density, canopy density, number of weeds, and precipitation during the flowering period. At present, there are 16 extended traits. The sowing date and emergence date are expressed in month-day. Lodging is divided into 5 grades according to the lodging rate. Grade 1 means not falling, and all plants stand

upright; Grade 2 means light fall, $0 < \text{lodging plant rate} \leq 25\%$; Grade 3 means mid-lodging, $25\% < \text{lodging plant rate} \leq 50\%$; Grade 4 refers to heavy lodging, $50\% < \text{lodging plant rate} \leq 75\%$; and Grade 5 refers to severe lodging, $\text{lodging plant rate} > 75\%$. Days of growth period refers to the total number of days from the emergence date to the mature period, expressed in days. The number of flowers refers to the total number of flowers on the flower axis. The purpose of expanding this trait is to establish a relationship with varieties. The rachis length is used to measure the number of flowers. Those with a short axis have fewer flowers and those with a long axis have more flowers. As shown in **Figure 2**, the pod shape is divided into straight shape, slightly curved sickle shape, and curved sickle shape, which are used to correlate with varieties and establish a yield prediction model. The mature color of the pod is the color of the pod when it is mature. It can be divided into five colors: Grass yellow, gray-brown, brown, dark brown, and black. The growth morphology can be divided into vine type, semi-erect type, and erect type according to the growth morphology of the main stem, which is used to establish the prediction model of unit yield. The color and spots of the leaf and the color and spots of the stem are used to establish the early warning model of diseases and pests. Both aforementioned extended traits can be represented in the form of pictures. The specific network can be trained



and tuned according to the acquired soybean leaf disease and insect image to achieve high leaf disease and pests' recognition accuracy. A convolution neural network was used to extract image features of four kinds of disease spots, and a support vector machine model for disease recognition was established (Qin et al., 2017). The retrieval and analysis of three soybean leaf diseases are realized by using the color, shape, and texture characteristics of images (Jayamala et al., 2016). Xu et al. (2020) uses the acquired small data samples of corn (*Zea mays*) leaf diseases and insect pests' images, a convolutional neural network based on transfer learning is proposed to identify corn diseases. The data set is expanded by data enhancement operations such as rotating and folding the original image. The model is improved based on the VGG-16 model. After comparing different migration learning training mechanisms and tuning the model, the average recognition accuracy of pests and diseases on the test set is as high as 95.33%. A trait feature descriptor-chord feature matrix is applied to the classification of soybean leaf images, which can distinguish the leaves of different soybean varieties, and is used to establish the relationship between varieties and leaf diseases and insect pests (Wang et al., 2017). Diseases can also develop on the stems of soybean plants. For example, soybean stem blight first occurs in the lower part of the stem, and then, gradually spreads to the upper part of the stem. In the early stage of the disease, the stems produce long oval lesions, which are grayish-brown, and then gradually expand into long black strips. After the leaves are fallen, the symptoms are more obvious on the stems of the plants before harvest, forming patches of oblong diseased spots. Therefore, images of soybean stems are collected to establish early warning models of diseases and insect pests, such as stem blight.

We use Dajiang UAV (DJI UAV) to capture the field distribution photographs on a sunny day and overcast and use CV technology to compare and process the acquired images of a sunny day and overcast, which can eliminate the influence of weather and obtain the traits of each plant such as plant height. These traits can be used to establish a yield forecast model. The DJI Matrice 600Pro UAV is used for field shooting, which is equipped with an RGB camera that is a digital camera, a near-infrared multi-spectral camera, and an infrared thermal imager to form the UAV imaging system (Zhou et al., 2020). They obtained field data of 116 soybean genotypes, and obtained 7 image features through image processing technology,

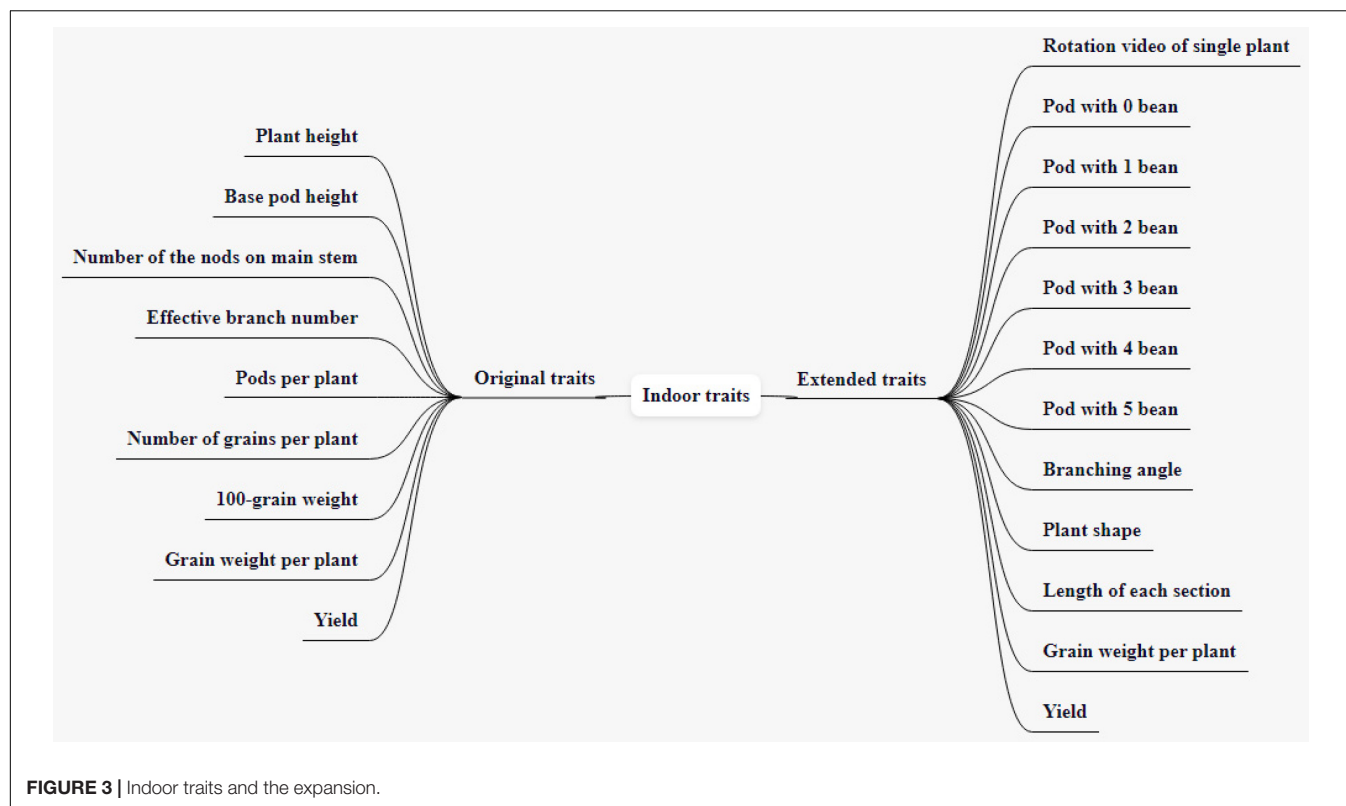
including temperature, hue, color saturation, canopy size, and plant height. A support vector machine model is developed, and the degree of soybean canopy withering is graded and is scored according to the acquired image features. Studies have shown that UAV-based image technology has great potential in selecting drought-tolerant soybeans. Therefore, obtaining field distribution photographs can extract the traits of each plant, which plays an important role in the subsequent establishment of a yield prediction model and the selection of drought-tolerant varieties. Field plant coverage is obtained by image processing on the original map of field distribution, which is used to establish relationships with other traits. The plant density is also obtained by processing field images, and the plant density is used to analyze the suitable density of soybean growth. Canopy density refers to the ratio of the total projected area of the canopy on the ground under direct sunlight (canopy width) to the total area of the ground, and it is used to reflect the density of planting. The number of weeds refers to the percentage of weeds in the number of plants, which is used to analyze the appropriate density of soybean growth.

The principle of obtaining the color of the leaf surface and the surface of the plant stem is the same as that of the flower color. As mentioned above, the colors and spots on the surface of leaves and plant stems are used to establish warning models of diseases and insect pests, so, the first thing to do is to mark the leaves with spots. The spots are divided into soybean downy mildew spots and soybean gray spot spots, soybean root rot spots, soybean brown spots, soybean bacterial spots, premature aging spots, etc. Then, use the marked image as a training set for the training of the target detection model. The purpose of target detection is to detect the spots in the new image and recognition.

Target detection technology, based on deep convolutional neural networks, has achieved good results in many fields (Zhang et al., 2019). The target detection method used in this study is YOLOV4 of the YOLO series. The network is mainly composed of four parts: input, Backbone, Neck, and prediction part. At the input end, Mosaic data enhancement methods are adopted to enrich the data set, which can adapt anchor calculation and image scaling. Mosaic data enhancement is a data enhancement method that takes four photographs and splices them in the way of random scaling, clipping, and arrangement. The backbone part refers to the convolutional neural network that aggregates features on different image granularity, while the neck part refers to a series of network layers with combined features, which finally transmits the feature map to the prediction layer, lastly, the head part refers to predicting the feature map, generating the prediction boxes, and giving the label.

QUANTITATIVE ANALYSIS OF SOYBEAN SEED TRAITS IN INDOOR TEST BASED ON COMPUTER VISION

Indoor traits include traits directly obtained through relevant simple operations indoors and traits obtained by 360° rotating video shooting of soybean plants fixedly harvested through specific devices, and processing pictures of some frames through



CV technology. Indoor traits include seven original traits and 12 expanded traits, as shown in **Figure 3**.

Original Indoor Traits

The original traits include plant height, base pod height, number of the nodes on the main stem, effective branch number, pods per plant, number of grains per plant, 100-grain weight, a total of seven traits. Plant height refers to the height from cotyledon node to plant top (including top inflorescence), expressed in cm, accurate to 0.1 cm, and the average value is taken for 10 consecutive plants. Base pod height refers to the height from the cotyledon node to the lowest pod insertion position, expressed in cm, the average value of 10 consecutive plants. The number of nodes on the main stem refers to the actual number of nodes at the top of the main stem (excluding the top inflorescence), which is related to the lodging resistance of the plant. The effective branch number refers to the number of branches with more than one pod, and the branches have at least two nodes. Excluding the secondary branches, take the average value of 10 consecutive plants. The number of pods per plant refers to the number of pods containing more than one full seed on a plant. The number of grains per plant refers to the number of grains obtained by one plant, including all intact grains, immature grains, insect eroded grains, and diseased grains. The 100-grain weight is obtained by randomly taking 100 grains from the intact grains of the sample 2 times, weighing them separately, and calculating the average value (if the difference between the two weighting values exceeds 0.5 g, resample and weigh), the unit is expressed in g. Soybean plants are fixed by a specific device, a soybean plant phenotype

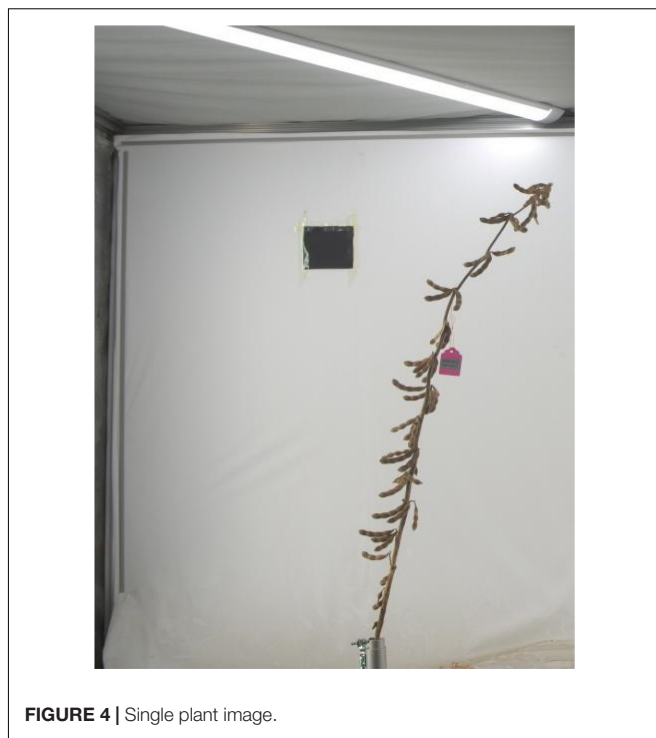
measuring instrument, and a background plate is installed to fix the shadow-absorbing screen to ensure the image quality. Indoor images of individual plants are obtained by the camera (BasleracA4112-20uc), with a resolution of 4,096px*3,000px. The camera parameter configuration is shown in **Table 1**. The background of the image is consistent with the lighting environment, and only color homogenization is required to deal with the slight background chromatic aberration. Both the plant height and the base pod height can be obtained by processing the image of the plant by CV technology. Compared with manual measurement, the error is within 1 cm. A fixed-length reference object is given next to the plant, and the plant and the reference object are taken together when the image is collected. We convert the RGB image to a grayscale image, set the grayscale value of the measurement target and the reference object to 1, and set the grayscale value of the rest to 0. We obtain the number of pixels of the target and reference, calculate the ratio of the number of pixels of the target to the number of pixels of the reference, and multiply the actual length of the reference with this ratio to get the actual length of the target. The number of the nodes on the main stem is obtained by calculating the number of joints of each node on the stem, and the gray image of the plant contour is obtained by edge detection technology. After eliminating the interference of non-branch contours, the effective branch number can be obtained by calculating the number of branch contours. The correct result rate of the number of the nodes on the main stem and effective branch number can reach 96%, which is slightly lower than that of naked eye recognition, but it greatly shortens the measurement time.

TABLE 1 | Basler acA4112-20uc camera parameters.

Parameters	Value
Pixel format	Bayer RG 8
Exposure auto	Off
Exposure time (μ s)	3,000
Gain auto	Off
Gain (dB)	0
Balance white auto	Continuous
Acquisition burst frame count	80
Trigger selector	Frame burst start
Trigger mode	On
Trigger delay (μ s)	500,000

Indoor Investigation Traits Extended by Computer Vision Techniques

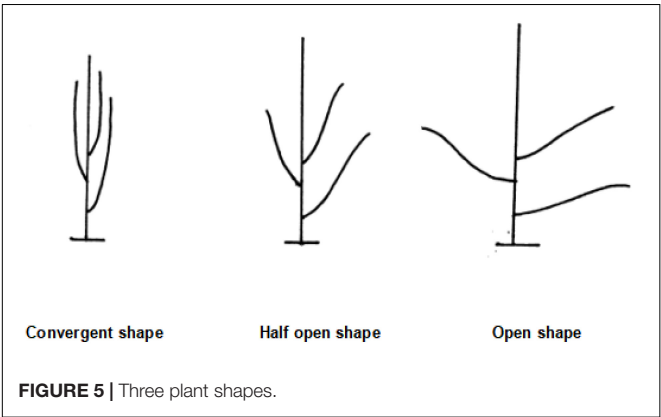
The expanded traits include grain weight per plant, yield, rotation video of a single plant, a pod with 0 beans, a pod with 1 bean, a pod with 2 beans, a pod with 3 beans, a pod with 4 beans, a pod with 5 beans, branching angle, plant shape, and length of each section, totaling 12 traits. The grain weight per plant refers to the grain weight of the sample (including immature, worm-eaten, and diseased grains) (g/plant). Yield is expressed in grams (g) and kept in whole numbers; the sun-dried seeds are weighed after the moisture content reaches below 13% and impurities are removed. Rotation video of a single plant refers to the video taken by a single plant rotating one circle on a mechanical device (soybean plant phenotype measuring instrument). The steps of the original soybean plant rotation video preprocessing method are as follows: to obtain each frame in the double plant rotation video, cut it into a single plant image by using the image clipping algorithm, then save it. The pixel of the single plant image is $2,048 * 3,000$. **Figure 4** shows the input single plant image. We capture images of certain frames in the video and use them as the original data for subsequent target detection. The background of these images is the same as the lighting environment, and no preprocessing is required. The 0–5 pods in the intercepted image were labeled, and appropriately improved the target detection algorithm YOLO V4 mentioned above. After the last layer of the model backbone network, the global attention module is introduced. The function of this module is the maximum pooling and average pooling of the global channel information are used to obtain new weights, and the re-weighting calculation generates the output of the module. The improved method extracts the protruding part of the pod and the edge of the pod as key features, which improves the ability of the traits of the model. After tuning and training the improved YOLO V4, the target detection model is obtained. The evaluation indicators of the model are shown in **Table 2**. The average accuracy of this model is 7% higher than that of YOLO V4, reaching the accuracy of manual identification, but the identification speed is up to 240 plants/hour, which is significantly improved compared with manual identification. The model can detect the pods on the plant and classify the pods belonging to several pods. All the images are input into the trained target detection model for detection, and the data of 5

**FIGURE 4** | Single plant image.

traits of 0–5 pods are obtained. According to the number of pods from 0 to 5, the number of pods per plant and the number of grains per plant can be calculated in common indoor copying traits. The number of pods from 0 to 5 is used to establish the relationship between variety and yield, and to establish a yield prediction model. Uzal et al. (2018) uses convolutional neural network CNN to classify pods with different numbers and confirms that CNN is significantly better than the classical method based on pre-designed feature extraction. The original image of this method must be a single-pod image of soybeans, and the Yolo model we used is less restrictive to the original image. It is not necessary to remove the pods from the plants to shoot, just select some frames of images from the video taken by rotating the whole plant 360 degrees. The branching angle, plant shape, and length of each section are all obtained by image processing. The branching angle refers to the angle between the main stem and the branch, which is used to judge the shape of the plant. According to the branch angle, the plant shape is divided into three types: convergent, semi-opened, and opened (**Figure 5**). The convergent shape refers to the small angle between the lower part of the branch and the main stem, and the upper and lower branches are compact; the open shape refers to the large branch angle, and the upper and lower branches are loose; the semi-open shape is between convergent and open. The length of each node and the number of main stem nodes in the original traits are obtained through image processing technology, with cm as the unit. The error between the calculated value of each section length and the actual measured value is within 0.2 cm. The recognition accuracy of the number of main stem nodes is as high as 98%. The branching angle, plant shape, and length of each section are used to establish contact with the variety.

TABLE 2 | Model evaluation index.

Model	Average precision (%)						Mean average precision (%)
	0 bean	1 bean	2 beans	3 beans	4 beans	5 beans	
Manual measurement	—	—	—	—	—	—	84
YOLOV4	79.40	85.81	90.37	96.36	97.20	18.51	77.94
Improved YOLOV4	89.26	98.00	99.83	98.26	98.98	21.91	84.37



Quantitative Analysis of Disease Resistance Traits Based on Computer Vision

The traditional investigation of disease resistance traits generally adopts the method of artificial inoculation identification, which requires a lot of manpower and material resources, and the cultivation cycle is long. How to combine technology with the prevention and control of diseases and insect pests is one of the important goals of achieving breeding. One of the goals. At present, there are more than 120 kinds of soybean pests and diseases reported in the world, and 52 kinds of soybean pests have been reported in my country. Based on the research background and existing experimental results, this study only focuses on the soybean diseases and pests that are more harmful in Northeast China. The original disease resistance traits are only gray spot disease and virus disease, so, it is expanded and increased on this basis.

According to the systematic investigation and research conducted by scientific researchers in 2015, the following diseases are common, and they are arranged according to the incidence degree of each disease (Zhu, 2014; Li et al., 2016; Liu, 2017; Gong, 2019).

Fungal Diseases

Types: (1) soybean root rot disease; (2) soybean downy mildew disease; (3) soybean brown stripe disease; (4) soybean brown spot disease; (5) soybean gray spot disease; (6) soybean stem blight disease; (7) soybean pseudo stem mildew stem blight disease; (8) soybean anthracnose disease; (9) soybean black spot disease; (10) soybean gray spot disease; (11) soybean damping-off; (12) soybean sclerotia disease; (13) soybean wheel blight; (14)

soybean vertical blight; (15) soybean sandwich blight; and (16) soybean purple spot.

Grades: Grade 0: the whole plant leaves are free of disease; Grade 1: partial disease, with less than five disease spots; Grade 2: a small number of diseased spots in the whole area, the distribution area of diseased spots accounted for less than 1/4 of the whole area; Grade 3: most plants in the whole region are infected, and the distribution area of disease spots accounted for half of the whole area; Grade 4: diseased spots are common in the whole area, and a few plants die early due to disease; and Grade 5: there are a lot of disease spots in plants in the whole area, and most plants die early due to disease.

Bacterial Diseases

Types : (1) soybean bacterial keratosis; and (2) soybean bacterial speckle.

Grade: Standard, refer to fungal diseases.

Viral Diseases

Types: Soybean Mosaic virus disease.

Grades: Grade 0: the leaves are flat in the whole area, without any signs of disease; Grade 1: 10% of plants in the whole region have 1–2 layers of upper leaf shrinkage; Grade 2: 20–40% of plants have withered upper leaves; Grade 3: more than 50% plants in the whole region have severe leaf shrinkage, yellow spots, affecting growth; and Grade 4: more than 70% of plants in the whole region have severe leaf shrinkage and bud dead branches.

Insect Pests

Types: (1) soybean cyst nematodes; (2) soybean heartworm; (3) soybean aphid; (4) soybean red spider; and (5) soybean meadow borer.

Grade: 0: the whole plant is free of disease; Grade 1: some leaves are infected, and the number of wormholes in infected leaves is less than 5; Level 2: a small number of plants in the whole region had wormholes in their leaves, and the distribution area of wormholes accounted for less than 1/4 of the leaf area; Level 3: most leaves of plants in the whole region have wormholes, and the distribution area of wormholes accounts for half of the leaf area; Level 4: wormholes are common in plant leaves, and a few leaves died early because of wormholes; and Level 5: there are many wormholes in plant leaves in the whole area, and most leaves die early because of wormholes.

Techniques can be used when investigating disease resistance traits. In the analysis stage, the target detection model can be established through computer image processing, and the area suffering from the disease can be successfully divided, and then,

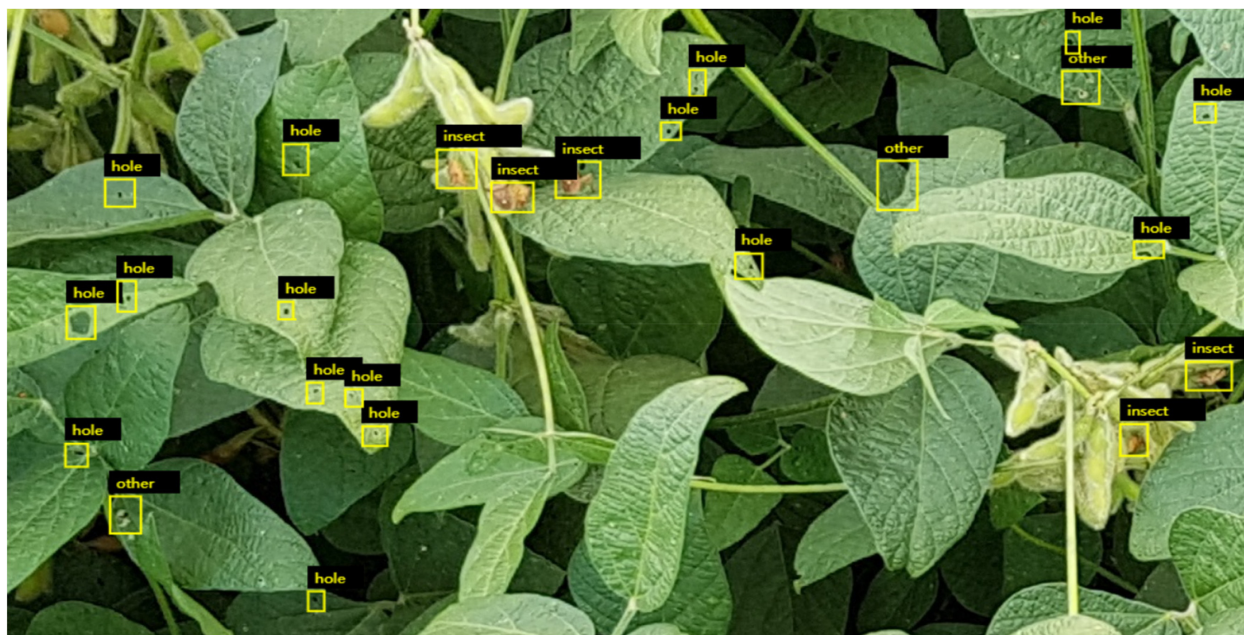


FIGURE 6 | Preliminary labeling pictures of leaf diseases and insect pests.

real-time monitoring should be conducted. The degree of disease can also be classified according to the results of image processing, and different prevention and control measures can be adopted according to different grades.

In this study, pictures of soybean in the field have been collected by UAV, and more than 2,000 pictures have been labeled to make data sets. The labeled data sets are shown in **Figure 6**. The target category was preliminarily divided into three categories, namely, wormhole, pest, and other disease spots. The YoloV5 model was trained for detection and recognition, and the recognition rates of all three categories reached 95%. Among them, the target objects identified as pests can be further identified in the subsequent stage, and other disease spots can also be further identified to analyze which kind of disease spots.

QUANTITATIVE ANALYSIS OF SEED TRAITS BASED ON COMPUTER VISION

The selection of soybean seeds is the most important step in seed examination. To obtain soybean seed phenotypes for machine seed selection by CV, we added soybean seed test traits as an important extension of the standard. We expanded the original soybean seed test traits. Soybean seed test traits include 10 original traits and 18 expanded traits, a total of 28 traits.

Original Soybean Test Traits

The original traits of the soybean test include seed shape, seed coat color quantity, seed coat color, seed coat spot type, cotyledon color, seed hilum color, seed coat cracking ratio, seed coat luster, seed crude protein content, seed crude fat content, a total of 10 traits, as shown in **Figure 7**. Seed shape is divided into five grades,

which are spherical, elliptic, long elliptic, oblong-elliptic, kidney-shaped. The number of seed coat colors can be divided into monochrome and bicolor. Seed coat color is divided into white, yellow, light yellow, yellow-green, green, light brown, brown, and black, a total of eight colors. The types of seed coat spots include tiger, saddle, and others. Midrib has three colors: yellow, chartreuse, and green. Umbilicus color is divided into light yellow, yellow, light brown, brown, light black, and black, a total of six colors. The seed coat cracking ratio is divided into nine grades: none or very low, none or very low to low, low to medium, medium to high, high to very high, and very high. Seed coat gloss is divided into none or none. The crude protein and crude fat content of seeds can be divided into 9 kinds: very low, very low to low, low, low to medium, medium, medium to high, high, high to very high, very high.

Soybean Test Traits Extended by Computer Vision Techniques

Soybean seed test traits are completely extended traits. Including seed number, male parent number, female parent number, area of hilum plane, the area of the marginal part of the hilum, the area inside the hilum, seed coat RGB, hilum margin position RGB, inside hilum RGB, seed coat glossiness, seed coat uniformity, photograph of hilum plane, photograph of kidney, photograph of hilum, length of hilum plane, width of hilum plane, girth of hilum plane, and degree of seed folding. There are 18 extended traits, as shown in **Figure 8**. It should be noted that the photographs of the hilum plane, kidney, and hilum may contain unextracted data, so, the three kinds of images are also represented as traits.

Seed number, male parent number, and female parent number are used to establish associations between phenotypic and

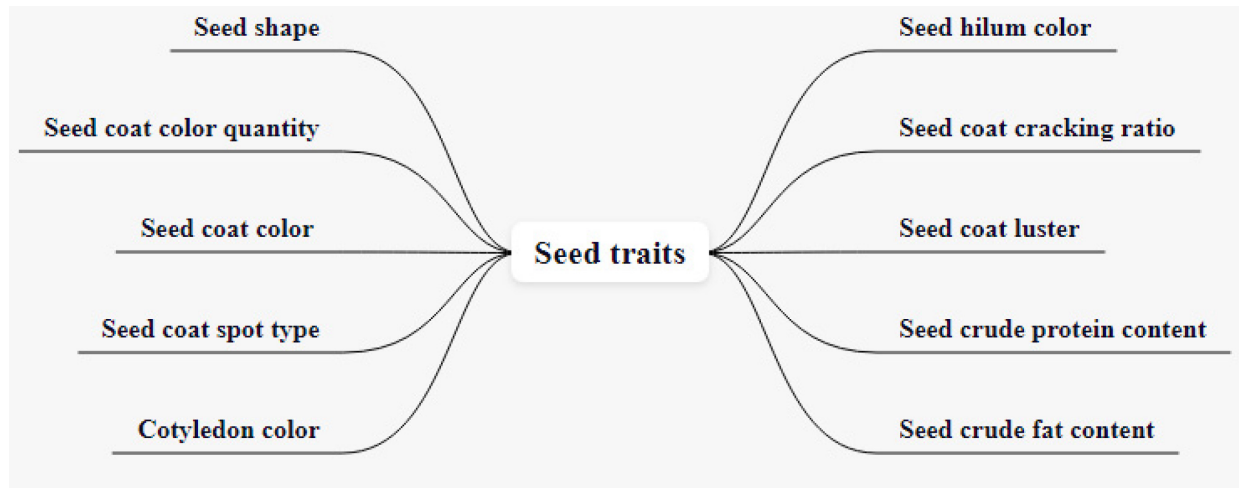


FIGURE 7 | Soybean test original traits.

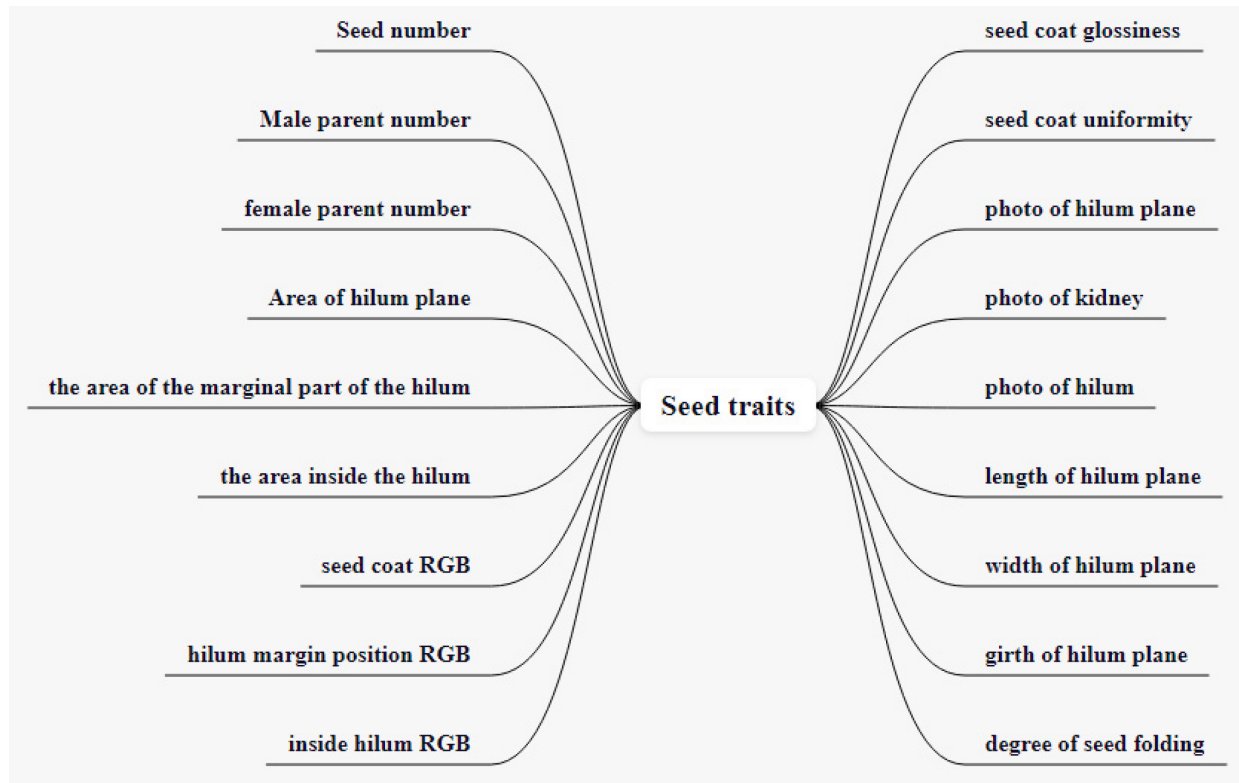


FIGURE 8 | Soybean test expanded traits.

genomic data and to establish a soybean knowledge graph. The original color traits are divided according to grades, and the expression is not accurate enough. So, we use RGB to represent color in extended traits. Seed coat RGB refers to seed coat color, used for color contrast between seed coats of different soybean varieties. The value range of seed coat glossiness is 0–1, used to describe the ability of seed coat to reflect light, replaced by the

V of HSV space. The seed coat uniformity ranges from 0 to 1 and is used to describe whether the brightness distribution of the seed coat is uniform from center to edge. The area of the hilum plane refers to the surface where the umbilicus is located and the area of the part without the umbilicus, in mm². The area of the marginal part of the hilum refers to the area of hilum, it includes the hilum margin (kidney), in mm². The area inside the

hilum refers to the surface on which the hilum is located; the area inside the hilum (hilum) is in mm^2 . Hilum margin position RGB refers to the color of the hilum margin (kidney) and is used to compare the color contrast of the hilum margin among different varieties. Inside hilum RGB refers to the color inside the hilum (hilum) and is used to compare the color inside the navel between different varieties. The photographs of the hilum plane are used to extract the features of the umbilicus. Photographs of the kidney is a seed hilum edge image extracted from the photographs of the hilum plane to extract seed kidney features. The photographs of the hilum are the internal images extracted from the photographs of the hilum plane and used to extract the traits of the hilum.

The hilum and kidney of seeds carry important genetic information. The area data of the above three parts can be used to establish the relationship between varieties and the prediction model of yield per unit area. The color traits of hilum are represented by RGB three-channel values and the values of a single channel range from 0 to 255. Values of different channels are separated by commas. The hilum and kidney of seeds carry important genetic information. The area data of the above three parts can be used to establish the relationship between varieties and the prediction model of yield per unit area. The color traits of hilum are represented by RGB three-channel values and the values of a single-channel range from 0 to 255. Values of different channels are separated by commas.

These traits, which carry genetic information, are delicate but important to breeders. The length, width, girth, and area of the hilum are used for comparison among different varieties, and the unit is expressed in pixels. **Figure 9** is a sample umbilical photograph from which we have extracted traits.

We conducted k-means clustering analysis on the data of the above four traits of 139,850 seeds extracted so far, which are divided into five categories as the standard, and the results are shown in **Figure 10**. **Figure 10** shows the corresponding quantities of five classes of the four traits. The five cluster centers of the length of the hilum plane are 90, 120, 146, 165, and 183, respectively, and it could be seen from the figure that most of the values of length of the hilum plane are clustered in the last three clusters. The five clustering centers of the width of the hilum plane are 89, 119, 144, 164, and 182, respectively, and most of the values are clustered in the latter four categories, that is, most of the values are more than 100. The five clustering centers of area are 6,602, 10,565, 15,656, 19,531, and 33,489, respectively, and most of the areas are clustered in the middle three categories, that is, the value is between 10,565 and 19,531. The five cluster centers of girth are 427, 600, 979, 2,811, and 5,553, respectively. Most of the values of the perimeter are clustered in the first two classes, indicating that most of the perimeter is less than 600. Through cluster analysis, the interpretability of data is increased. Hierarchical representation of the data of the four traits is more beneficial to establish the relationship between varieties and the prediction model of yield per unit area.

The HSV model is a color model similar to the way the human eye perceives color, which can intuitively express the light and shade, tone, and brightness of colors, facilitating color comparison and emotional expression. In this study, the lightness V in the HSV model is used to measure the seed coat glossiness.

The conversion formula of the image from RGB color space to HSV color space is as follows:

$$\begin{cases} R' = R/255 \\ G' = G/255 \\ B' = B/255 \end{cases} \quad \begin{cases} \max = \max(R, G, B) \\ \min = \min(R, G, B) \end{cases} \quad (1)$$

$$H = \begin{cases} 0^\circ, & \text{if } \max = \min \\ 60^\circ \times \left(\frac{G' - B'}{\max - \min} + 0 \right), & \text{if } \max = R' \\ 60^\circ \times \left(\frac{B' - R'}{\max - \min} + 2 \right), & \text{if } \max = G' \\ 60^\circ \times \left(\frac{R' - G'}{\max - \min} + 4 \right), & \text{if } \max = B' \end{cases} \quad (2)$$

$$S = \begin{cases} 0, & \max = 0 \\ \frac{\max - \min}{\max}, & \max \neq 0 \end{cases} \quad (3)$$

$$V = \max \quad (4)$$

The mean lightness of the whole seed region is calculated in the transformed HSV space as the final evaluation result of soybean seed glossiness. Seed coat uniformity reflects the uniformity of skin color distribution and is an important indicator to measure the quality of seeds. The method adopted in this study is to calculate the standard deviation δ from the RGB value of skin color in the whole seed region to represent the dispersion of seed skin color distribution. The formula is as follows:

$$\delta = \sqrt{\frac{\sum_{x=0}^{l-1} \sum_{y=0}^{b-1} [P(x, y) - \mu]^2}{A}} \quad (5)$$

The L refers to the pixel length of soybean seed area, B refers to the pixel width of soybean seed area, A refers to the pixel area of soybean seed, P refers to the RGB value at the midpoint (X, Y) of seed image, and μ refers to the RGB mean value of seed coat area. The value of seed uniformity is between 0 and 1, and the closer the value is to 0, the more uniform the seed coat color distribution is. Due to the irregular traits of soybean seeds, it is difficult to measure the long and short axes directly. The length of the outer rectangle is the long axis of the soybean seed, and the width is the short axis of the soybean seed. According to the coordinate values of the four vertices of the outer rectangle, the specific pixel values of the long and short axes of soybean seeds are calculated. The girth of the hilum plane is obtained by calculating the number of pixels contained in the contour extracted from the edge of the seed image. The hilum is a strip on the concave side of soybean seed. It is a long, thin scar left after the seed has fallen off the seed stalk or placenta. It is round, oval, oval, etc. The color and shape of seed hilum of different soybean varieties are significantly different, so, the phenotypic traits of seed hilum could be used as an important parameter for seed identification. Seed kidney refers to the kidney-shaped region around the seed hilum, which is relatively dark in color and fuzzy in edge when observed by naked eyes. According to the experience of breeders, different varieties of soybean have significant differences in the performance of this part, so, the seed kidney is also used as

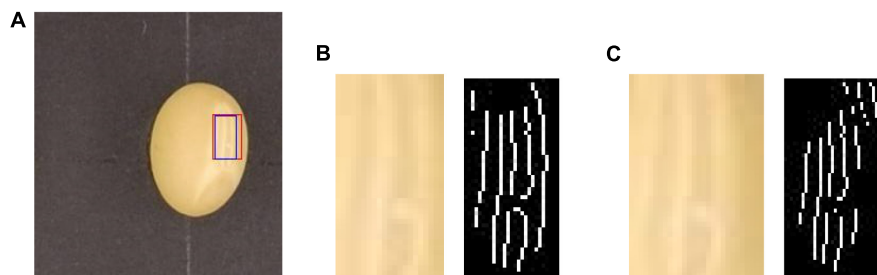


FIGURE 9 | Photograph of a soybean seed (A) and umbilicus (B) and seed kidney (C) obtained using CV techniques.

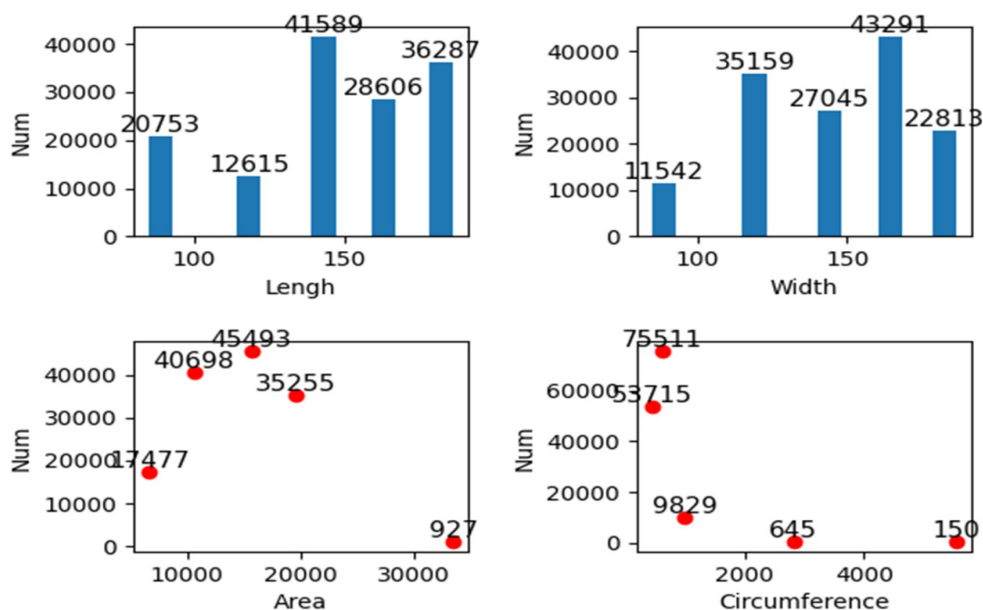


FIGURE 10 | Cluster analysis of length, width, perimeter, and area.

an important phenotypic traits parameter to identify soybean seeds. The area of the hilum and kidney is very small, only relying on observation and measurement of the naked eye can't be accurate and precise. With the aid of CV technology, we proposed an algorithm for identifying hilum and kidney based on edge detection, then, finally achieved the contour detection and segmentation of this narrow part and extracted the phenotypic traits parameters of this key part, which provided a new data reference index for soybean seed test and breeding program design. Gaussian Laplacian (LoG) operator is an edge detection operator based on the second-order differential method, which is composed of Gaussian filter and Laplace operator. The Laplace operator is used to extract the edge of the image by second-order differential operation, which can reflect the minimal change of gray value well and is suitable for edge detection with the small difference of gray value between seed hilum, seed kidney, and seed coat of soybean. In general, before edge detection, a two-dimensional Gaussian function is needed to denoise the image to improve the robustness of the operator to noise and discrete points, and then, edge detection is carried out. The point in the

image whose second derivative is 0 is the target edge point. The two-dimensional Gaussian function can be expressed as follows:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (6)$$

Let $I(x, y)$ be the input image, then the calculation formula of edge detection by LoG operator is as follows:

$$\hat{I}(x, y) = \nabla^2 G(x, y) I(x, y) \quad (7)$$

When using the LoG operator to detect the edge of the hilum and kidney, there will be an edge breakpoint. Therefore, a method based on the breadth-first search is proposed to optimize the edge detection result of the LoG operator. The algorithm first classifies the detected target edge pixels to determine whether it is an endpoint; the method is to search the eight neighborhood pixels around the pixel. If only one of the eight neighborhood pixels is connected to the pixel, the pixel is an endpoint, and the endpoint is saved in the endpoint array, otherwise, the pixel is a non-endpoint. For the non-endpoint, if there is no pixel

3	2	1
4	(x, y)	0
5	6	7

FIGURE 11 | Pixel eight neighborhoods graph.

connected to it in the neighborhood, it is an isolated point. A point is on the edge-line if two endpoints are connected to it in a neighborhood of eight; If there are four points connected to it in the neighborhood, it is an edge-line crossing point. The eight-neighborhood map of pixels is shown in **Figure 11**. Then, the breadth-first algorithm is used to search the pixels in the endpoint array until the search stop condition is met, and the search path is recorded. Then, the endpoint with the shortest Euclidian distance is selected from the breakpoint to connect, and the complete hilum and kidney edges are obtained. In the actual process, it is necessary to screen the edge segments searched, remove the interference segments that do not belong to the seed hilum and seed kidney parts, then, connect the edge segments that meet the requirements, and finally, obtain the complete edge contour of soybean seed hilum and seed kidney. The specific steps of the algorithm are shown in **Figure 12**.

DATABASE AND SOYBEAN VISUAL ARTIFICIAL INTELLIGENCE BREEDING PLATFORM

Systematic collection, management, and analysis of phenotypic data have shown certain genetic gains in accelerated breeding (Araus et al., 2018). Therefore, the construction of a phenotypic database is an essential step. The database entity relationship (ER) diagram corresponding to various original traits is shown in **Figure 13**, and the database ER diagram corresponding to expanded traits is shown in **Figure 14**.

The database structure and relationship corresponding to original traits are shown in **Figure 13**, which only included field traits table (tjxz), disease traits table (disease), and indoor traits table (snxz). The field traits table includes variety number (bean_id), flowering period (khq), colors of flowers (hs), shapes of leaf (yx), fluffy color (rms), density of trichome (rmmd), number of lobules (xys), green degree of leaves (green_degree), deciduous (lyx), growth form (szxt), pod bending degree (wq_degree), blasting pods (zjx), pod color (p_color), pod bearing habit (jjxx), and mature period (csq), with a total of 15 fields. The disease traits table contains three fields, including variety number, gray spot (hbb), and virus disease (bdb). The indoor traits table included plant number (zhizhu_id), variety number (bean_id), plant height (height), base pod height (dijia_height), number of the nodes on main stem (zjjs), effective

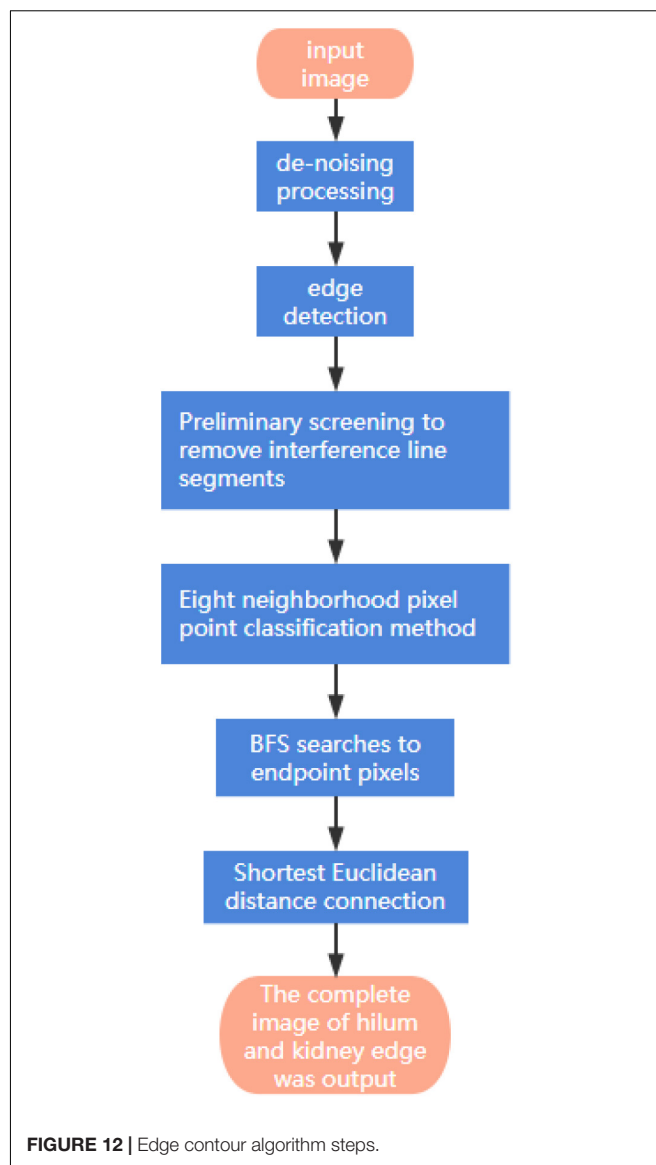


FIGURE 12 | Edge contour algorithm steps.

branch number (yxfzs), pods per plant (dzjs), number of grains per plant (dzls), and 100-grain weight (blz), with a total of 9 fields. The soybean seed traits table (zzxz) includes seed number (zz_id), variety number (bean_id), seed shape (zz_shape), seed coat color quantity (yssl), seed coat color (zpys), seed coat spot type (sblx), cotyledon color (zyys), seed hilum color (zqys), seed coat cracking ratio (klbl), seed coat luster (zpgz), seed crude protein content (cdb), and seed crude fat content (czf). In the four tables, the plant number (zhizhu_id) of the indoor traits table (snxz) is used as both a primary key and a foreign key, which is associated with the seed number (zz_id) of the seed traits table. The variety number (bean_id) is used as the foreign key to be associated with the primary key variety number (bean_id) of the field traits table (tjxz). The variety number (bean_id) of the disease traits table (disease) is both a primary key and a foreign key, which is associated with the various number (bean_id) of the field traits table. The seed traits table (zzxz) is a table, in which the

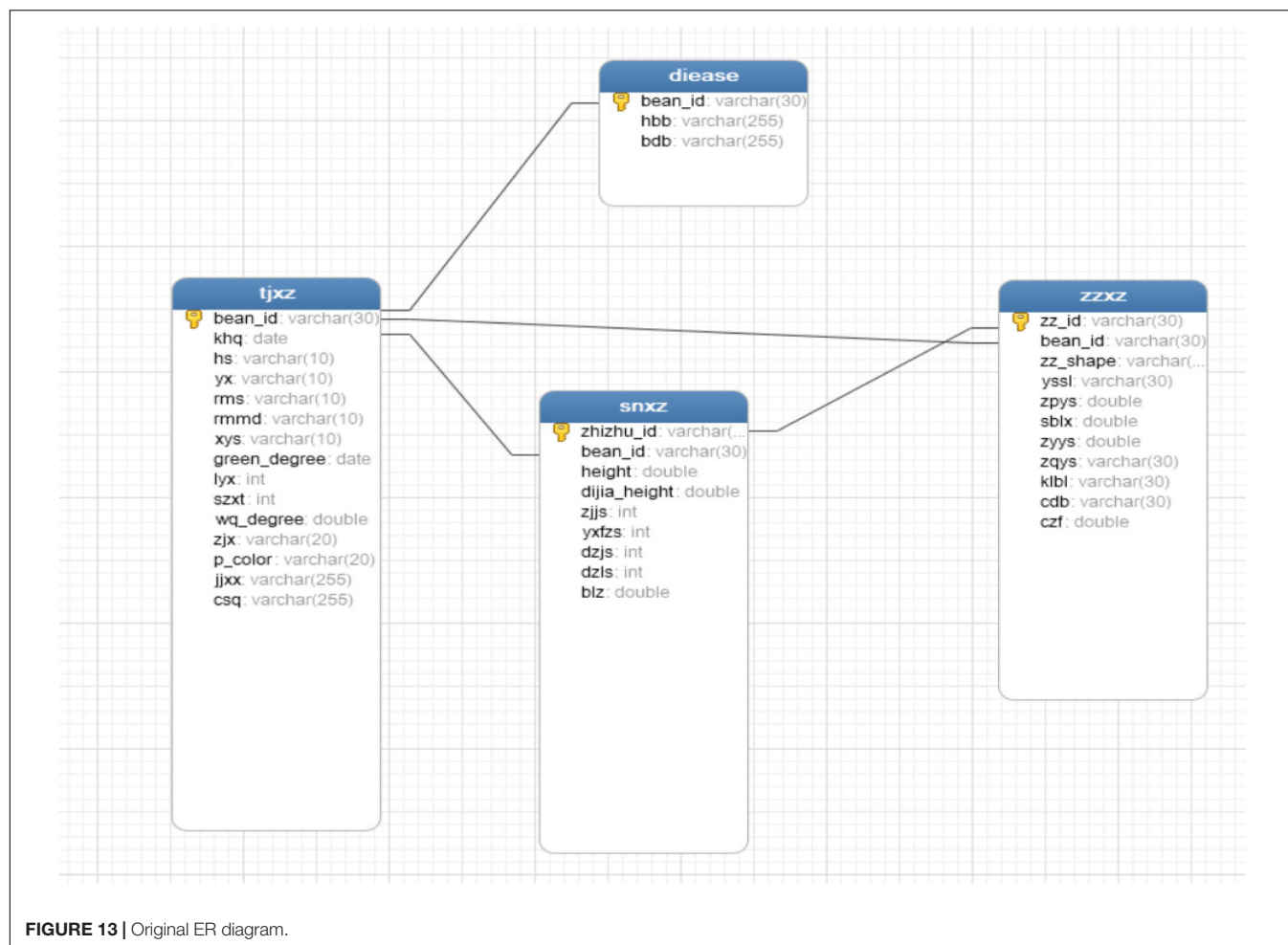


FIGURE 13 | Original ER diagram.

seed number (*zz_id*) acts as the primary key, the variety number (*bean_id*) serves as a foreign key in the table and is associated with the primary key in the field traits table.

The structure and relationship of database tables corresponding to existing traits are shown in **Figure 14**, including field traits table (*tjxz*), disease traits table (*disease*), indoor traits table (*snxz*), and seed traits table (*zzxz*). The expanded field traits table (*tjxz*) includes variety number (*bean_id*), flowering period (*khq*), colors of flowers (*hs*), shapes of leaf (*yx*), fluffy color (*rms*), density of trichome (*rmmd*), number of lobules (*xys*), green degree of leaves (*green_degree*), deciduous (*lyx*), growth form (*szxt*), pod bending degree (*wq_degree*), blasting pods (*zjx*), pod color (*p_color*), pod bearing habit (*jjxx*), mature period (*csq*), sowing date (*bzq*), emergence date (*cmq*), lodging (*dfx*), days of growth period (*syrs*), number of flowers (*hds*), rachis length (*hzcd*), pod shape (*jgxx*), color and spots of leaf (*yp_col_spots*), color and spots of stem (*j_col_spots*), field distribution photographs (sunny day) (*s_pho*), field distribution photographs (overcast) (*c_pho*), field plant coverage (*fgl*), canopy density (*ybd*), amount of weeds (*zcl*), precipitation during flowering period (*kh_jsl*), and plant density (density). The expanded indoor traits table (*snxz*) includes plant number (*zhizhu_id*), variety number (*bean_id*), plant height

(*height*), base pod height (*dijia_height*), number of the nodes on main stem (*zjjs*), effective branch number (*yxfzs*), pods per plant (*dzjs*), number of grains per plant (*dzls*), 100-grain weight (*blz*), grain weight per plant (*dzlz*), yield (*cl*), rotation video of single plant (*sp*), pod with 0 bean (*0lj*), pod with 1 bean (*1lj*), pod with 2 beans (*2lj*), pod with 3 beans (*3lj*), pod with 4 beans (*4lj*), pod with 5 beans (*5lj*), branching angle (*fzjd*), plant shape (*zx*), and length of each section (*mjcd*). The expanded disease traits table (*disease*) includes variety number (*bean_id*), fungal disease (*zjl*), bacterial disease (*xjl*), viral disease (*bdx*), and insect pest (*ch*). The expanded soybean seed traits table (*zzxz*) includes seed number (*zz_id*), variety number (*bean_id*), seed shape (*zzxz*), seed coat color quantity (*yssl*), seed coat color (*zpys*), seed coat spot type (*sblx*), cotyledon color (*zyys*), seed hilum color (*zqys*), seed coat cracking ratio (*klbl*), seed coat luster (*zpgz*), seed crude protein content (*cdb*), seed crude fat content (*czf*), male parent number (*father_id*), female parent number (*mother_id*), area of hilum plane (*zqmmj*), the area of the marginal part of the hilum (*zqbymj*), the area inside the hilum (*zqnbmj*), seed coat RGB (*zpRgb*), hilum margin position RGB (*zqbyRgb*), inside hilum RGB (*zqnbRgb*), seed coat glossiness (*zp_gzd*), seed coat uniformity (*zp_jyd*), photograph of hilum plane (*zqm_img*), photograph of kidney (*zs_img*), photograph of hilum (*zq_img*),

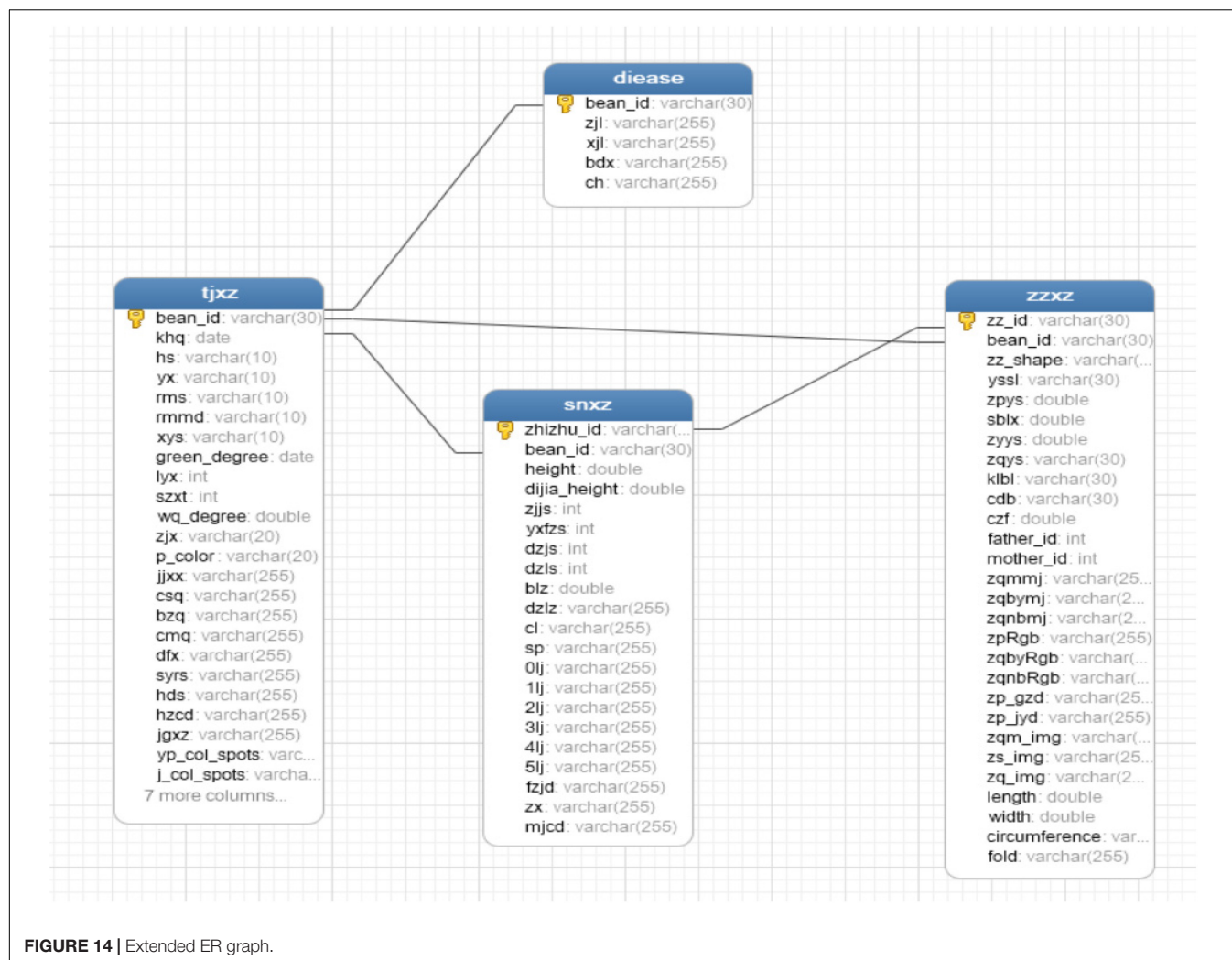


FIGURE 14 | Extended ER graph.

length of hilum plane (length), width of hilum plane (width), girth of hilum plane (circumference), and degree of seed folding (fold). The relationship between the four tables after expansion is the same as before.

In the original table structure, there is a lack of phenotypic trait information, and the disease table could only store the information of two diseases, and the correlation between the four tables is low. In the expanded table structure, the gaps of seed phenotypic traits are filled in, and the disease traits table could store four kinds of disease information, enrich the fields of each table, and increase the correlation among the tables. Among them, if the plant number in the indoor traits table is the same as the seed number in the seed traits table, it means that the phenotypic information of the plant is the corresponding phenotypic information of the seed when it grows into a plant, thus, associating the information of two stages of the life of the same soybean.

Firstly, the field information of each variety is obtained through mobile phone terminal and CV technology, and the field information of each variety is stored in the field traits table. Then, parts of the collected soybean disease images are used for

deep learning training, so that the trained model can recognize each disease. The model is used to identify the disease images of different varieties, and the disease image information and disease category of different varieties are saved in the disease traits table. Variety numbers and various information of all varieties are included in the field trait table. The variety number in the disease traits table is restricted by the variety number in the field traits table. Before planting, the soybean is fixed on a specific device for image collection, and the data of various phenotypic traits of seeds are obtained by using computer image processing technology and CV technology, and all seed data are imported into the seed traits table in batches. After the plants corresponding to the seeds are harvested, the plants are fixed on a specific rotating device in the indoor space for 360° video collection. Images of specific frames are captured in the video to label pods with each grain number, and then, the improved YOLOV4 algorithm mentioned above is used to train the model. The trained model could recognize several pods, and the number of pods, pods per plant, and seeds per plant could also be obtained. The phenotypic information of base pod height and plant height is also obtained by CV technology after capturing

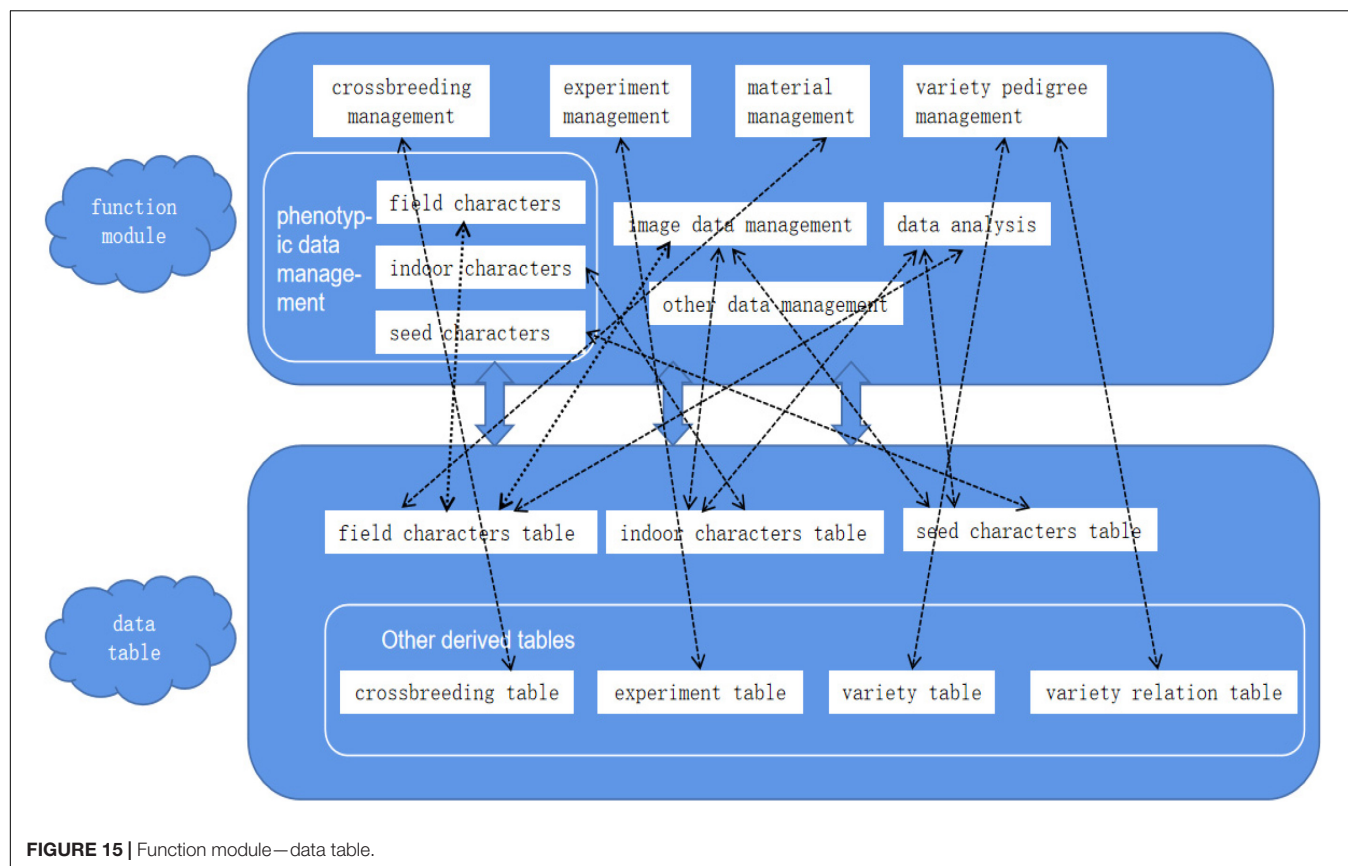


FIGURE 15 | Function module—data table.

specific frames of images. Finally, the obtained plant information is imported into the indoor traits table in batches. The phenotypic information of a seed is associated with the corresponding plant information by numbering. Variety numbers of indoor and seed traits table are constrained by variety numbers of field traits table.

The soybean visual AI breeding platform corresponding to the database system is divided into 8 functional modules, which are crossbreeding management, experiment management, material management, variety pedigree management, phenotypic data management, image data management, data analysis, and other data management. Phenotypic data management is divided into field traits, indoor traits, and seed traits. The data tables in the ER diagram above and derived data tables together constitute a complete database system, and each table is closely linked with functional modules, as shown in **Figure 15**. The close connection of the data table, and the forming of the functional correlation, so that the phenotypic information of the seed, the phenotypic information of the plant corresponding to the seed, and the phenotypic information of the variety to which the seed belongs can be queried at a time, bringing the previously independent loose data closer together. Some fields in each table can also be input into the SVM model as predictors for training, and the final model can predict the category of a certain trait. The filtering of fields as predictors and the determination of target fields are the focus of subsequent work. When the amount of data reaches a certain level, big data technology can be used to mine and analyze the data of each field.

CONCLUSION

In this study, three types of traits including field traits, indoor test traits, and disease traits are expanded. At the same time, soybean seed traits are added, and the role of the expanded phenotypic traits is explained. It is convenient to use CV technology to quickly obtain data and then, perform machine learning, which can provide big data support for breeding experts to design accurate breeding programs.

The CV test has a greater advantage over the traditional manual test and the earlier CV test. For the extraction of phenotypic traits such as plant height and base pod height, the error is within one pixel, that is, the accuracy reaches millimeter level. The image processing program processes images in batches. Compared with the traditional manual test, the data collection speed is greatly improved, and it has high throughput. The area of each part of the extracted soybean seeds and the accuracy of the length and width data of the seeds have also reached the millimeter level. These phenotypic data cannot be obtained by traditional artificial seed testing.

The improved YoloV4 algorithm training model is used to detect the pods of plants. At present, the average recognition accuracy of various types of pods is about 84.37%, and the average recognition accuracy of one to four pods has reached more than 90%. The model is still being further tuned and tuned. There are large number of pods on a single soybean plant and many pods shade each other. Manual observation will inevitably

cause errors and consume a lot of time. This method is more accurate and efficient to obtain the data of the number of each pod and completed tasks that are difficult for traditional manual test species. Traditional plant phenotypic information acquisition mainly relies on manual measurement, naked eye observation, and software analysis after image acquisition. The grain size of soybean seeds is small and contains multiple phenotypic traits, so, it is difficult to obtain phenotypic data by manual measurement and naked eye observation, and some image analysis software is not flexible enough to meet the needs of breeders to obtain all the required phenotypic traits. By using CV and machine learning technology, soybean phenotypic traits can be acquired more accurately and efficiently. Compared with traditional methods, phenotypic traits are obtained more comprehensively. This method is a supplement to current plant Phenomic methods. Traditional methods can be used to validate the data acquired by machine learning. The phenotypic traits of a soybean are the physical expression of its genetic genes. More data of phenotypic traits are extracted, which could be used to establish the prediction model of soybean yield per unit area, and to provide data support for establishing the relationship between phenotypic traits and genome.

All traits of soybean directly or indirectly affect breeding. Many studies have shown that some phenotypic traits directly affect soybean yield. The phenotypic traits of a soybean are all intuitive, visible, and easy to measure, but the deeper level of the soybean gene remains to be studied (Cheng et al., 2016). Propose a method for rapid extraction of soybean seed DNA with a success rate of more than 98%, which can be used for genetic analysis. With the further application of

technology in seed tests, CV technology can obtain more or more accurate traits, which can provide more big data support for breeding experts to optimize breeding programs. The problem we will solve is not only to establish the relationship between phenotypic traits and genomes, and the relationship between soil and ecological environment of crop growth, but also to establish a soybean big data platform, providing breeding experts with big data services in an easy-to-understand way of knowledge expression.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

YX contributed to the research of data expansion and the writing of the article. PL contributed to the pest part. HH contributed to the conception and modification of the article. JL and HY provided guidance for professional knowledge. XF contributed to the modification of the article. All authors contributed to the article and approved the submitted version.

FUNDING

The research was supported by Hainan Yazhou Bay Seed Lab (No. B21HJ0101).

REFERENCES

- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., and Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.* 23, 451–466. doi: 10.1016/j.plants.2018.02.001
- Cheng, W., Zhengjun, X., Xianzhong, F., and Suxin, Y. (2016). Establishment and application of a rapid and non-destructive soybean seed DNA extraction method. *Acta Bot.* 51, 68–73.
- Gong, X. (2019). Analysis on the prevention and control of common soybean diseases and insect pests. *Agric. Dev. Equip.* 10, 180–182.
- Guo, R., Chongyu, Y., Hong, H., et al. (2021). Detection method of soybean pod number per plant using improved YOLOv4 algorithm. *Trans. Chin. Soc. Agric. Eng.* 37, 179–187.
- Hu, W., Fu, X., Chen, F., and Yang, W. (2019). The development road of a new generation of plant phenomics. *Acta Bot.* 54, 558–568.
- Jayamala, K. P., Raj, K. and Asian Agricultural and Biological Engineering Association. (2016). Analysis of content based image retrieval for plant leaf diseases using color, shape and texture features. *Eng. Agric. Environ. Food* 10, 69–78. doi: 10.1016/j.eaef.2016.11.004
- Li, M., Wang, Y., Chen, J., Yan, X., Liu, X., Duan, Y., et al. (2016). Investigation on the types and damage levels of soybean fields in Northeast China in 2015. *Soybean Sci.* 35, 643–648.
- Liu, G. (2017). In 2015, the types of soybean fields in Northeast China and the degree of damage are clear. *Pest. Mark. Inf.* 5:63.
- Ma, H. (2020). *The Collection and Analysis of Soybean Seed Phenotypic Characteristic Data for Artificial Intelligence Breeding*. Jinan: Shangdong University.
- Ministry of Agriculture and Rural Affairs of the People's Republic of China (2018). Guidelines for Testing Plant Variety Specificity, Consistency and Stability of Soybean. GB / T 19557.4-2018. 2018-05-14. Available online at: <https://www.chinesestandard.net/PDF/English.aspx/GBT19557.24-2018> (accessed May 05, 2018).
- Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., and Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Sci.* 282, 60–72. doi: 10.1016/j.plantsci.2019.01.007
- Pound, M. P., Atkinson, J. A., Townsend, A. J., Wilson, M. H., Griffiths, M., Jackson, A. S., et al. (2018). Deep Machine Learning provides state-of-the-art performance in image-based plant phenotyping. *Gigascience* 6, 1–10. doi: 10.1093/gigascience/gix083
- Qin, F., Liu, D., Sun, B., Ruan, L., Ma, Z., and Wang, H. (2017). Image recognition of four different alfalfa leaf diseases based on deep learning and support vector machine. *J. China Agric. Univ.* 22, 123–133.
- Ren, S., Jia, F., Gu, X., et al. (2020). Recognition and segmentation model of tomato leaf diseases based on deconvolution-guiding. *Trans. Chin. Soc. Agric. Eng.* 36, 186–195.
- Uzal, L. C., Grinblat, G. L., Namías, R., Larese, M. G., Bianchi, J. S., Morandi, E. N., et al. (2018). Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agric.* 150, 196–204. doi: 10.1016/j.compag.2018.04.024
- Wang, B., Chen, L., and Ye, M. (2017). String feature matrix: an effective shape descriptor for plant leaf image classification and retrieval. *Chin. J. Comput.* 40, 2559–2574.

- Xu, J., Shao, M., Wang, Y., and Han, W. (2020). Convolutional neural network image recognition of corn diseases based on transfer learning. *Trans. Chin. Soc. Agric. Mach.* 51, 230–236.
- Zhang, S., Gong, Y., and Wang, J. (2019). The development of deep convolutional neural networks and their applications in the field of computer vision. *Chin. J. Comput.* 42, 453–482.
- Zhou, J., Francois, T., Tony, P., John, D., Daniel, R., Neil, H., et al. (2018). Plant phenomics: history, present status and challenges. *J. Nanjing Agric. Univ.* 41, 580–588. doi: 10.7685/jnau.201805100
- Zhou, J., Zhou, J., Ye, H., Ali, M. L., Nguyen, H. T., and Chen, P. (2020). Classification of soybean leaf wilting due to drought stress using UAV-based imagery. *Comput. Electron. Agric.* 175:105576. doi: 10.1016/j.compag.2020.105576
- Zhu, Z. (2014). Prevention and control methods of common soybean diseases and insect pests in Northeast China. *Jilin Agric.* 23:71.

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 Xing, Lv, He, Leng, Yu and Feng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) 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.



Rapid Identification of Soybean Varieties by Terahertz Frequency-Domain Spectroscopy and Grey Wolf Optimizer-Support Vector Machine

Xiao Wei^{1,2†}, Dandan Kong^{1†}, Shiping Zhu^{2*}, Song Li², Shengling Zhou² and Weiji Wu³

¹ College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou, China, ² College of Engineering and Technology, Southwest University, Chongqing, China, ³ China Tianjin Grain and Oil Wholesale Trade Market, Tianjin, China

OPEN ACCESS

Edited by:

Xianzhong Feng,
Northeast Institute of Geography
and Agroecology, Chinese Academy
of Sciences (CAS), China

Reviewed by:

Hong He,
Shandong University, China
Xiaohui Yuan,
Wuhan University of Technology,
China
Bin Wang,
Nanjing University of Finance
and Economics, China

*Correspondence:

Shiping Zhu
zspswu@126.com

[†] These authors have contributed
equally to this work

Specialty section:

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

Received: 15 December 2021

Accepted: 17 January 2022

Published: 11 March 2022

Citation:

Wei X, Kong D, Zhu S, Li S,
Zhou S and Wu W (2022) Rapid
Identification of Soybean Varieties by
Terahertz Frequency-Domain
Spectroscopy and Grey Wolf
Optimizer-Support Vector Machine.
Front. Plant Sci. 13:823865.
doi: 10.3389/fpls.2022.823865

Different soybean varieties vary greatly in their nutritional value and composition. Screening for superior varieties is also essential for the development of the soybean seed industry. The objective of the paper was to analyze the feasibility of terahertz (THz) frequency-domain spectroscopy and chemometrics for soybean variety identification. Meanwhile, a grey wolf optimizer-support vector machine (GWO-SVM) soybean variety identification model was proposed. Firstly, the THz frequency-domain spectra of experimental samples (6 varieties, 270 in total) were collected. Principal component analysis (PCA) was used to analyze the THz spectra. After that, 203 samples from the calibration set were used to establish a soybean variety identification model. Finally, 67 samples from the test set were used for prediction validation. The experimental results demonstrated that THz frequency-domain spectroscopy combined with GWO-SVM could quickly and accurately identify soybean varieties. Compared with discriminant partial least squares (DPLS) and particles swarm optimization support vector machine, GWO-SVM combined with the second derivative could establish a better soybean variety identification model. The overall correct identification rate of its prediction set was 97.01%.

Keywords: soybean, DPLS, PSO-SVM, GWO-SVM, THz spectroscopy

INTRODUCTION

Soybean is one of the most important raw materials for oil and feed (Herman et al., 2018; Kumar et al., 2021; Wei et al., 2021a). Differences in soybean varieties lead to significant differences in their protein, fat, and other constituent contents (Wang et al., 2020a,b). At the same time, soybean variety screening has a crucial impact on the quality of soybean products. Currently, common methods for soybean variety identification include simple sequence repeat (SSR) molecular marker assays (Lu et al., 2018; Wen et al., 2020) and detection of soybean components to determine their varieties (Larsen, 1967; Ujiie et al., 2005), among others. Although the accuracy of the above methods is relatively high, and the sensitivity is relatively strong, and the application is relatively wide. However, they have problems such as relatively long time consuming, relatively low efficiency,

and relatively complicated detection process. In recent years, Near-infrared spectroscopy (NIRS) technology has been introduced for the detection of agricultural varieties (Lun Liu et al., 2010; Teye et al., 2014). Compared to SSR molecular marker assays and the soybean component-based detection variety method, the NIRS technology has the advantage of not requiring pre-treatment of samples. Nevertheless, it has limitations in detecting soybeans with surface defects (Zhu et al., 2010) and limited detection accuracy (Chen et al., 2019; Rong et al., 2020). Hence, it is essential to study a rapid and accurate identification method suitable for different varieties of soybeans.

Terahertz (THz) spectroscopy has unique advantages in soybean variety identification (Wei et al., 2020, 2021b). THz spectroscopy is based on coherent THz pulses generated by ultrafast optics. It is a broadband linear spectral detection technique. Due to the weak interaction forces between biological macromolecules (hydrogen bonding, van der Waals forces), backbone vibrations and dipole rotations, etc. fall right in the THz spectral range. At the same time, THz pulses have a good temporal resolution (on the order of picoseconds). Therefore, THz spectroscopy technology is currently cross-cutting frontier research that is received great attention (Yang et al., 2016). Currently, there has been some research on the identification of agricultural product and food varieties through THz spectroscopy. For instance, Wu et al. (2020) proposed a method for sesame oil variety identification based on THz time-domain spectroscopy. Eventually, the identification model using radial basis kernel function achieved a 100% identification rate. Yang et al. (2021) used THz spectroscopy and competitive adaptive reweighted sampling (CARS) combined with support vector machine (SVM) for the detection of high oil and common maize. Ultimately, the model identification rate could reach 100%. Ge et al. (2015) applied THz spectra and partial least squares regression (PLSR) models to discriminate wheat varieties. Eventually, the prediction accuracy of the optimized model using interval partial least squares was significantly improved. The related coefficient of prediction set for their wheat variety detection model was 0.992. Li et al. (2019) used THz spectra and a neural network learning vector quantization model for qualitative identification of maize varieties. By changing the ratio of dividing the training and prediction sets, the final prediction set had a 100% discrimination rate. Luo et al. (2019) conducted a study on soybean variety identification using THz spectroscopy and integrated learning methods. The studied pre-processing methods, integrated classifiers, and comparison methods. Finally, the average accuracy of the proposed model was 89.29%. In summary, there are relatively few reports on soybean variety identification based on THz frequency-domain spectroscopy, and such related studies still have some academic value and significance.

The objective of the study was to analyze the feasibility of THz frequency-domain spectroscopy and chemometrics to identify soybean varieties. Also, a soybean variety identification model based on the grey wolf optimizer-support vector machine (GWO-SVM) was proposed. After different pre-processing methods, the discrimination results of three [discriminant partial least squares (DPLS), particles swarm optimization-support vector machine

(PSO-SVM), and GWO-SVM] soybean variety identification models were compared. The most appropriate pre-processing method for each variety identification model was selected.

MATERIALS AND METHODS

Experimental Materials

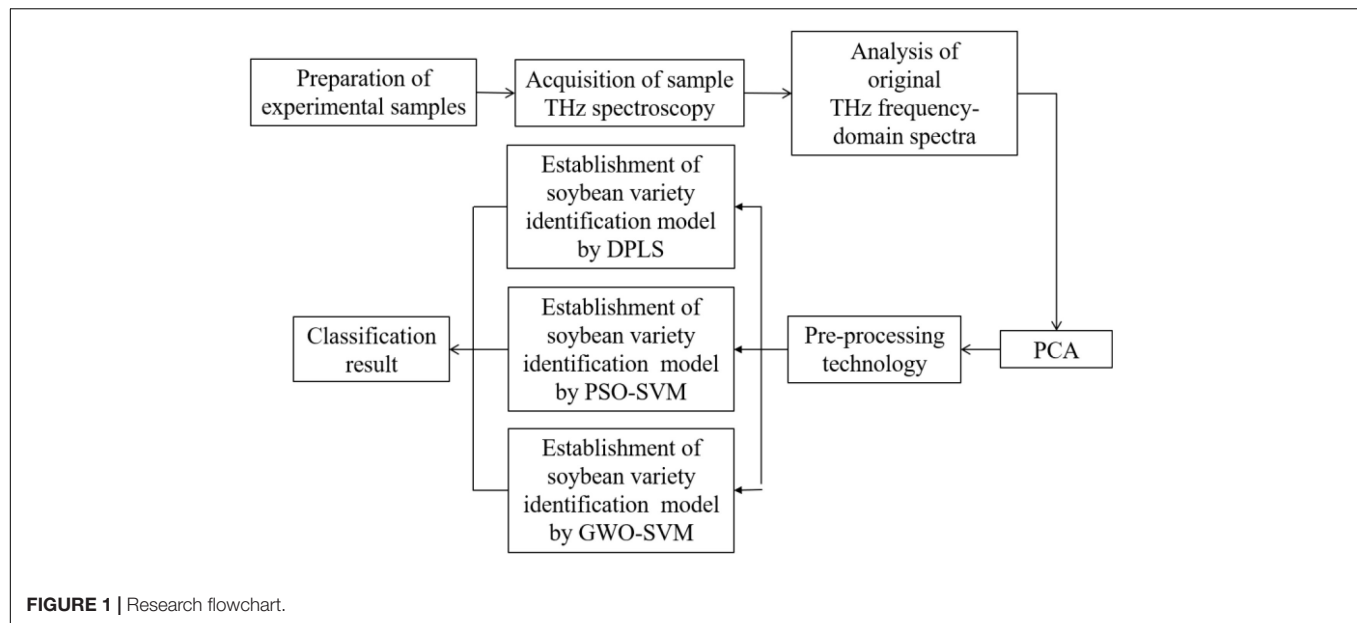
Soybean samples of six varieties (HuaiDou 2, LuDou 1, NiuMoHang, LuDou 4, HeDou 12, QiHang 34, abbreviated as HD₂, LD₁, NMH, LD₄, HD₁₂, QH₃₄) were collected for this experiment. Among them, HD₂, LD₁, and NMH were each two batches. LD₄, HD₁₂, and QH₃₄ were each three batches. Each batch of soybean samples was weighed 50 g to perform subsequent experiments. Eighteen experimental samples were prepared for each batch of soybean samples. A total of 270 samples were prepared. The soybean samples used in the experiments were collected by the Quality Inspection Center of the Tianjin Grain and Oil Wholesale Trading Market in China. The quality inspection center conducted the soybean sample collection in strict accordance with soybean varieties. This made the subsequent soybean variety identification experiments in this paper more rigorous and accurate.

Terahertz Spectroscopy Experimental Equipment

In the experiment, the THz spectroscopy equipment from EKSPLA was used for the spectral data acquisition of the experimental sample. The equipment used the FF50 femtosecond laser as the ultrashort pulse laser source. The central wavelength was 1,064 nm, and the pulse duration was less than 150 fs. The repetition frequency was about 80 MHz, and the output power was greater than 40 mW, and the spot diameter was less than 2 mm. The equipment used low-temperature-grown gallium arsenide as the generator and detector of the THz wave. The optical distance between the generator and detector was about 62.5 cm. The pump light source was divided into two beams of 55:45 by the beam splitter after passing through the half-wave plate. The first pump light was guided by the reflector through the fast delay line. After that, it was then directed through a set of the optical lens into the THz emitter to excite the THz pulse. The second part of the pump laser beam passed through the slow delay line. Afterward, it was guided to the THz detector by a reflector. The THz pulse was incident vertically on the experimental sample through a metal parabolic mirror. Then, it was focused to reach the THz detector, where it converged with the second part of the beam. The beam signal was fed to the lock-in amplifier for amplification. Finally, The THz time-domain spectra of the experimental samples were obtained.

Experimental Sample Preparation and Terahertz Spectrum Acquisition

Firstly, the soybean samples were dried in a 40°C drying oven for 3 h. This reduced the moisture in the samples during transportation and storage, thus reducing the effect of moisture in the soybean samples on the experiment. Afterward, the



soybean samples were crushed using a pulverizer. The crushed samples were then further ground through a mortar and pestle to obtain the soybean sample powder. Secondly, the soybean sample powder was filtered through the sieve with pore sizes of 0.074 mm. Later, the filtered sample powder was taken and added to polyethylene powder (sample powder and polyethylene powder were mixed in the ratio of 7:3). The two powders were mixed thoroughly to obtain the experimental sample powder. Finally, the experimental sample powder was weighed 135 mg using a precision balance. The sample powder was pressed under the pressure of 20 MPa to form a flake with a thickness of about 1 mm. The surface of the flakes was ensured to be smooth. The room temperature of the THz spectroscopy acquisition laboratory was controlled at 25°C. Nitrogen gas was charged into the THz spectroscopy experimental equipment before the start of the experiment. The relative humidity in the experimental equipment was kept below 5% at all times. During the experiment, the experimental samples were loaded into the sample holder and their THz time-domain spectra were scanned. Each experimental sample was scanned 256 times, and a total of 6 sample points were scanned. The THz spectra of the six sample points were averaged. The THz time-domain spectra were converted to THz frequency-domain spectra by the device software. The THz spectra were acquired by the software that came with the THz spectroscopy experimental equipment.

The main research flow chart of this paper is shown in **Figure 1**.

Theory

Principal Component Analysis

Principal Component Analysis (PCA) is a common way of data analysis. In order to extract the main features and information of the THz spectra of experimental samples, PCA is often performed on the spectral data (Wang et al., 2022). The main purpose

of PCA is to reduce the dimensionality of THz spectral data as a way to exclude the numerous chemical information that overlaps each other. It mainly highlights the similarities and differences of the data. This is because data identification is difficult to achieve in high-dimensional data. It uses new variables to represent the original variables. These new variables do not lose useful information in the original variables as much as possible (Rezazad Bari et al., 2021). The new variables are called Principal Components (PC).

Pre-processing Methods

Experimentally acquired THz spectra often contain some interferences from factors unrelated to the nature of the sample. These interferences can cause baseline drift of the spectrum and generate random noise, etc. At the same time, the absorption peaks often appear to overlap. Therefore, it is essential for the spectral data to be subjected to pre-processing methods. The pre-processing method can amplify the original hidden signal differences in the spectral data. Meanwhile, spectral pre-processing techniques can achieve the purpose of improving the resolution of THz spectral data, making the identification more accurate and reliable (Ndlovu et al., 2021; Tafintseva et al., 2021). In this paper, seven pre-processing methods were used, including: mean-centering, auto scaling, standard normal variate (SNV), minimum and maximum values to [0 1], multiplicative scatter correction (MSC), first derivative, and second derivative (Lu, 2006; Chu, 2011).

Discriminant Partial Least Squares

Discriminant Partial Least Squares (DPLS) is a discriminant analysis method based on PLSR (Lei et al., 2021). It is a widely used method for supervised pattern discrimination. This method considers the experimental sample characteristics data as the independent variables X (whose rows are the sample ordinal numbers and columns are the characteristic variable ordinal

numbers). The category information of the experimental samples is considered as the dependent variable Y . Y is a matrix composed of 0, 1. The rows correspond to the sample numbers. The columns correspond to the category serial numbers. When a sample belongs to a category, the element value of the corresponding column in Y is 1. Otherwise, it is 0. In order to decide the problem of attribution of a substance in a mixture, the category matrix must be able to describe a specific kind of sample (Xue et al., 2021). DPLS is commonly applied in cases where the number of variables is high and there are multiple commonalities.

Particles Swarm Optimization-Support Vector Machine

SVM is a very widely used pattern recognition model proposed based on statistical theory (Wang et al., 2021b). In this paper, radial basis function was used as the kernel function of SVM. The classification hyperplane established by SVM can guarantee the classification accuracy (Wang et al., 2021a). For the optimization problem of the parameters of the SVM (parameter c and g), this paper used the particles swarm optimization (PSO) algorithm (Huang et al., 2021) and the grey wolf optimizer (GWO) algorithm (Deng et al., 2021). PSO is an optimization algorithm for group intelligence. It is derived from the study of predatory behavior of birds. The basic idea of the PSO algorithm is to find the optimal solution through collaboration and information sharing among individuals in a population (Zhou et al., 2021b). Each particle in this algorithm represents a potential solution to the problem. The velocity of the particle is dynamically adjusted with the movement experience of itself and other particles, thus achieving individual optimality search in the solvable space.

Grey Wolf Optimizer-Support Vector Machine

GWO is a meta-heuristic optimization algorithm. It has a more reasonable global optimal solution search mechanism, greater operational stability, and faster convergence than other optimization algorithms (Liu et al., 2021). The GWO algorithm is proposed based on imitating the hunting process of a wolf pack. It is mainly divided into three steps, which are encirclement, hunting, and attack. The highest rank in this algorithm is the head wolf, with two of them, marked as α . The head wolf is responsible for making decisions and leading the pack during the hunting (finding the optimal parameters) process. The remaining wolves are, respectively, labeled β , δ , and ω from top to bottom according to rank. The behavior of the next rank follows the leadership of the previous rank. Firstly, the wolves encircle the target during the hunting process. After encircling the prey, the wolves perform hunting behavior. The process is usually led by α , β , and δ . Other search units (ω) should update their respective positions according to the current position of the best search unit. Finally, the wolves attack the prey and accomplish the goal of capturing the prey (Zhou et al., 2021a).

DPLS, PSO-SVM, and GWO-SVM soybean variety identification models were established and predicted done in MATLAB R2018a. The computer operating system was Windows 10.0. The CPU was i7 8750 H. The memory is 16 g 2,666. In this paper, the correct identification rate was calculated in the same way as the accuracy.

RESULTS AND DISCUSSION

Terahertz Frequency-Domain Spectra

Figure 2 shows the THz frequency-domain spectral images of the experimental samples in the interval of 0.1–1.5 THz and 0.1–2.5 THz. The effective range of THz spectra measured by the THz spectroscopy equipment used in this experiment was from 0.1 to 2.5 THz. When the THz spectral frequency was between 1.5 and 2.5 THz, the signal-to-noise ratio of the spectrum was too low to be selected. Hence, the 0.1–1.5 THz interval was selected as the modeling spectral interval for the soybean variety identification model in this experiment. It could not be seen from Figure 2 that there were significant differences in the THz spectra of the different variety experiment samples. Therefore, the THz spectra of the experimental samples should be analyzed and identified in combination with chemometrics.

Principal Component Analysis

The THz frequency-domain spectra of the samples were subjected to PCA. The cumulative variance contribution of the first 3 PCs was 99.65%. Therefore, the information of the distribution characteristics of the samples could be basically characterized by the projected distribution of the first 3 PCs in space. The PC score plot is shown in Figure 3. From Figure 3, it could be found that the experimental samples of HD₂ and LD₁ were distributed more scattered in the three-dimensional space. However, the remaining four varieties of experimental samples showed obvious overlap in the distribution in three-dimensional space. In the overlapping part, it was very difficult to distinguish and identify the experimental sample varieties using the naked eye. Therefore, good results could not be obtained by using PCA alone to identify soybean varieties. Thus, THz spectroscopy required the use of identification methods with supervised modes for soybean variety identification.

Variety Identification Model

Establishment and Validation

Establishment and Validation of the Discriminant Partial Least Squares Soybean Variety Identification Model

The transmissibility and frequency of the spectral points in the selected frequency interval were used as the input matrix. The DPLS was used to establish the soybean variety identification model. The 270 experimental samples were divided according to the ratio of calibration set to test set of 3:1. Therefore, 67 samples were randomly selected as the test set. The remaining 203 samples were used as the calibration set. The 67 test set samples contained 10 HD₂, 10 LD₁, 7 NMH, 13 LD₄, 16 HD₁₂, and 11 QH₃₄. The 203 calibration set samples contained 26 HD₂, 26 LD₁, 29 NMH, 41 LD₄, 38 HD₁₂, and 43 QH₃₄. Firstly, the THz spectra were separately subjected to seven pre-processing methods. Secondly, the DPLS soybean variety identification model was established by the calibration set. Finally, the effects of the variety identification model were validated using the test set. The validation results are shown in Table 1.

From Table 1, it could be found that the results of the DPLS soybean variety identification model were not very satisfactory.

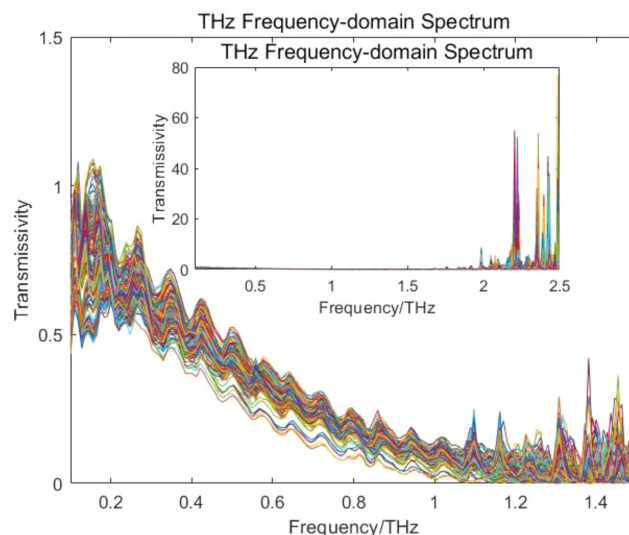


FIGURE 2 | THz frequency-domain spectra of different variety experimental samples.

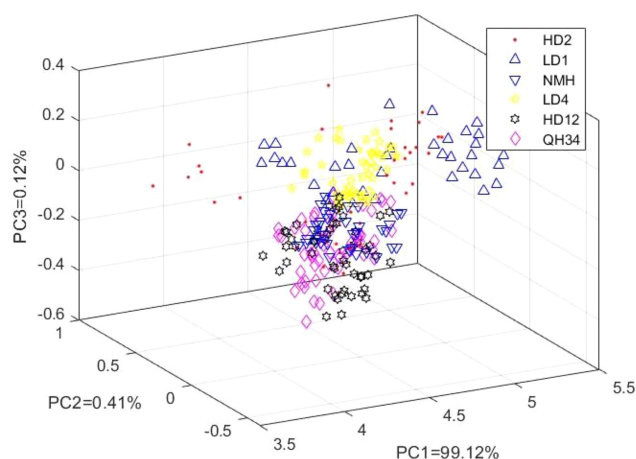


FIGURE 3 | The PC score plot.

The overall correct identification rate was in the range of 74–81%. The validation results of the DPLS soybean variety identification model showed relatively obvious changes after the THz frequency-domain spectra were subjected to different pre-processing methods. In terms of the overall correct identification rate, the overall correct identification rate of the DPLS soybean variety identification model could be improved to 80.60% after the second derivative pre-processing of the THz spectra. This was a 4.48% improvement compared to the DPLS soybean variety identification model without the pre-processing method, and the identification time used was relatively shorter. At the same time, the overall precision of this identification model achieved the highest value. This might be because the second derivative pre-processing method could effectively eliminate the interference of baseline and other backgrounds. Judging from the identification of different variety soybeans, the DPLS soybean

variety identification model was very effective in identifying LD₄ and QH₃₄. The DPLS soybean variety identification model was significantly improved for NMH by the second derivative pre-processing. The DPLS soybean variety identification model was not well for the identification of three soybean varieties (HD₂, LD₁, and HD₁₂). Compared to other varieties, the identification results of these three varieties of soybean needed to be further improved.

Establishment and Validation of the Particles Swarm Optimization-Support Vector Machine and Grey Wolf Optimizer-Support Vector Machine Soybean Variety Identification Models

The experimental samples were divided into the calibration and test set according to the same method as before. The calibration set was formed by 203 experimental samples. The PSO-SVM

TABLE 1 | DPLS soybean variety identification model validation results.

Spectral pre-processing methods	The best number of PC	Correct identification rate%					Overall precision%	Overall F1 score%	Identification time used s
		HD ₂	LD ₁	NMH	LD ₄	HD ₁₂	QH ₃₄		
(a). None	9	40	60	85.71	100	68.75	100	75.64	0.90
(b). Mean-centering	10	40	80	85.71	100	68.75	100	78.64	0.51
(c). Auto scaling	10	50	60	85.71	100	75	100	78.66	0.35
(d). SNV	9	20	70	85.71	100	68.75	100	72.88	0.47
(e). Minimum and maximum values to [0 1]	10	50	50	85.71	100	81.25	100	78.40	0.47
(f). MSC	9	20	70	85.71	100	68.75	100	72.88	0.79
(g). First derivative	10	50	70	85.71	100	56.25	100	82.13	0.65
(h). Second derivative	9	50	70	100	100	68.75	100	80.42	0.62

The row with the highest value of overall correct identification rate% and precision% is highlighted in bold.

and GWO-SVM soybean variety discrimination models were established using the calibration set. The test set was composed of 63 experimental samples. Validation of the soybean variety discrimination models was performed using the test set. The parameters c and g were, respectively, optimized by the PSO algorithm and GWO algorithm. The validation results of the two soybean variety identification models are shown in **Tables 2, 3**.

From **Tables 2, 3**, it could be found that after the first derivative and the second derivative pre-processing methods, the overall correct identification rate of the GWO-SVM soybean variety identification model was very significantly improved compared to the PSO-SVM variety identification model. The identification time used was also significantly reduced. Meanwhile, the overall precision of the GWO-SVM variety identification model achieved the highest value. This might be because the parameter optimization of the SVM by GWO imitated the wolf hunting process so that it could obtain a more reasonable global optimal solution search capability. Therefore, the GWO algorithm showed superior soybean variety identification performance compared to the PSO algorithm in finding the optimal parameters of the SVM. This was of great practical importance for soybean variety identification. The overall correct identification rate of the GWO-SVM soybean variety identification model ($c = 7.77 \times 10^9$, $g = 7.95 \times 10^{-4}$) was improved to 97.01% after the second derivative pre-processing method for the THz frequency-domain spectra. The identification time used of the variety identification model was 181.66 s. This indicated that THz frequency-domain spectroscopy combined with chemometrics could quickly and accurately identify soybean varieties. After THz frequency-domain spectra were preprocessed with the second derivative, the GWO-SVM variety identification model improved the overall correct identification rate by 7.46% compared to the GWO-SVM identification model without the pre-processing method. This further indicated that the second derivative pre-processing method played an important role in eliminating background interference, resolving overlapping peaks, etc. for the THz spectra. Thus, the second derivative pre-processing method best improved the overall correct identification rate of the GWO-SVM identification model. After the second derivative pre-processing, the GWO-SVM identification model could reach 100% for four varieties (HD₂, LD₁, LD₄, and QH₃₄) of soybeans. The correct identification rate of the other two varieties of soybeans also reached more than 85%. By observing the validation results of the two variety identification models combined with the seven pre-processing methods, it was easy to see that the THz spectra combined with different pre-processing methods had a great impact on the correct identification rate of the identification model and the time used for identification. Therefore, it was crucial to choose the appropriate pre-processing method for different identification models.

Comparing **Tables 1–3**, it was found that the overall correct identification rate of the GWO-SVM soybean variety identification model was better than that of the DPLS and PSO-SVM variety identification models after the first derivative and second derivative preprocessing methods. However, the DPLS soybean variety identification model took significantly shorter time to identify than the other two variety identification models.

TABLE 2 | PSO-SVM soybean variety identification validation results.

Spectral pre-processing methods	Correct identification rate%							Overall precision%	Overall F1 score%	Identification time used s
	HD ₂	LD ₁	NMH	LD ₄	HD ₁₂	QH ₃₄	Overall			
None	100	90	100	92.31	75	100	91.04	93.35	91.29	196.04
(b). Mean-centering	100	90	100	92.31	75	100	91.04	93.35	91.29	230.47
(c). Auto scaling	100	90	100	100	75	100	92.54	94.84	92.77	331.03
(d). SNV	90	90	85.71	100	75	100	89.55	91.11	89.81	239.65
(e). Minimum and maximum values to [0 1]	100	80	100	100	75	100	91.04	93.71	91.25	278.69
(f). MSC	90	90	85.71	100	75	100	89.55	91.11	89.81	152.84
(g). First derivative	90	80	100	92.31	81.25	100	89.55	90.90	89.72	211.92
(h). Second derivative	100	80	71.43	100	87.5	100	91.04	91.44	90.99	248.08

The row with the highest value of overall correct identification rate% and precision% is highlighted in bold.

TABLE 3 | GWO-SVM soybean variety identification validation results.

Spectral pre-processing methods	Correct identification rate%							Overall precision%	Overall F1 score%	Identification time used s
	HD ₂	LD ₁	NMH	LD ₄	HD ₁₂	QH ₃₄	Overall			
(a). None	100	90	100	84.62	75	100	89.55	92.13	89.81	162.33
(b). Mean-centering	100	90	100	84.62	75	100	89.55	92.13	89.81	147.19
(c). Auto scaling	100	90	100	100	75	100	92.54	94.84	92.77	330.74
(d). SNV	90	100	85.71	92.31	75	100	89.55	91.50	89.89	218.83
(e). Minimum and maximum values to [0 1]	100	80	100	100	75	100	91.04	93.71	91.25	322.81
(f). MSC	90	100	85.71	92.31	75	100	89.55	91.50	89.89	182.52
(g). First derivative	100	100	100	92.31	81.25	100	94.03	95.51	94.20	160.37
(h). Second derivative	100	100	85.71	100	93.75	100	97.01	97.01	97.01	181.66

The row with the highest value of overall correct identification rate% and precision% is highlighted in bold.

When comparing the validation results of the DPLS, PSO-SVM, and GWO-SVM identification models, it was found that the THz spectra combined with the second derivative pre-processing method resulted in the best identification results and relatively short identification time for the GWO-SVM variety identification model. The overall correct identification rate was 97.01% (85.71% for NMH, 93.75% for HD₁₂, and 100% for others), and the identification time used was 181.66 s. However, there were some limitations of this variety identification model. The identification model needed to be continuously optimized for different THz spectral data. For soybean varieties for which THz spectral data characteristics had been not collected, the identification model identified relatively poor results.

CONCLUSION

The experimental results showed that it was feasible to identify soybean varieties by THz frequency-domain spectroscopy combined with chemometrics. The GWO-SVM soybean variety identification model achieved the best results and relatively short identification time used after the second derivative pre-processing method for THz spectra. The overall correct identification rate was 97.01% and the identification time used was 181.66 s. This indicated that this method was an accurate means of identifying soybean varieties. The identification time used was relatively short, and the identification speed was relatively fast. In addition, the DPLS and PSO-SVM variety

identification models combined with suitable pre-processing methods could also be used for soybean variety identification. The novelty of the paper was that the feasibility of THz frequency-domain spectroscopy combined with chemometrics for soybean variety identification was analyzed and investigated. At the same time, a soybean variety identification model based on the GWO-SVM was proposed. The study has some reference value for the rapid and accurate identification of agricultural products and food varieties based on THz spectroscopy.

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

XW: conceptualization, methodology, software, validation, formal analysis, investigation, writing—review and editing, writing—original draft, and visualization. DdK: investigation, resources, data curation, writing—review and editing, and supervision. SpZ: definition, validation, resources, data curation, writing—review and editing, project, administration, and funding. SL: formal analysis and investigation. SLZ: resources, data curation, and administration. WjW: investigation, resources,

and data curation. All authors contributed to the article and approved the submitted version.

FUNDING

The research was supported by the National Natural Science Foundation of China (Grant Nos. 31801257, 62005227, and 31771670), the Chongqing Natural Science Foundation (Grant No. cstc2020jcyj-msmX0300), and the Graduate Scientific

Research and Innovation Project of Chongqing (Grant No. CYB20099).

ACKNOWLEDGMENTS

We want to thank Fengxu Li and Yongfa Li for their precious support in sample acquisition and preparation. We further thank Wanqin Zheng for her helpful discussions on the manuscript.

REFERENCES

- Chen, J., Li, M., Pan, T., Pang, L., Yao, L., and Zhang, J. (2019). Rapid and non-destructive analysis for the identification of multi-grain rice seeds with near-infrared spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 219, 179–185. doi: 10.1016/j.saa.2019.03.105
- Chu, X. (2011). *Molecular Spectroscopy Analytical Technology Combined with Chemometrics and its Applications*. Beijing: Chemical Industry Press.
- Deng, J., Chen, W. L., Liang, C., Wang, W. F., Xiao, Y., Wang, C. P., et al. (2021). Correction model for CO detection in the coal combustion loss process in mines based on GWO-SVM. *J. Loss Prev. Process Ind.* 71:104439. doi: 10.1016/j.jlp.2021.104439
- Ge, H., Jiang, Y., Lian, F., Zhang, Y., and Xia, S. (2015). Characterization of wheat varieties using terahertz time-domain spectroscopy. *Sensors* 15, 12560–12572. doi: 10.3390/s150612560
- Herman, R. A., Ekmay, R. D., Schafer, B. W., Song, P., Fast, B. J., Papineni, S., et al. (2018). Food and feed safety of DAS-44406-6 herbicide-tolerant soybean. *Regul. Toxicol. Pharmacol.* 94, 70–74. doi: 10.1016/j.yrtph.2018.01.016
- Huang, W., Liu, H., Zhang, Y., Mi, R., Tong, C., Xiao, W., et al. (2021). Railway dangerous goods transportation system risk identification: comparisons among SVM, PSO-SVM, GA-SVM and GS-SVM. *Appl. Soft Comput.* 109:107541. doi: 10.1016/j.asoc.2021.107541
- Kumar, V., Vats, S., Kumawat, S., Bisht, A., Bhatt, V., Shivaraj, S. M., et al. (2021). Omics advances and integrative approaches for the simultaneous improvement of seed oil and protein content in soybean (*Glycine max* L.). *Crit. Rev. Plant Sci.* 40, 398–421. doi: 10.1080/07352689.2021.1954778
- Larsen, A. L. (1967). Electrophoretic differences in seed proteins among varieties of soybean, *Glycine max* (L.) Merrill 1. *Crop Sci.* 7, 311–313. doi: 10.2135/cropsci1967.0011183x000700040008x
- Lei, L., Ke, C., Xiao, K., Qu, L., Lin, X., Zhan, X., et al. (2021). Identification of different bran-fried *Atractylodes Rhizoma* and prediction of atractylodin content based on multivariate data mining combined with intelligent color recognition and near-infrared spectroscopy. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 262:120119. doi: 10.1016/j.saa.2021.120119
- Li, H., Wu, J., Liu, C., Sun, X., and Yu, L. (2019). Research on classification and identification of maize varieties based on LVQ and THz time domain spectra. *J. Chinese Cereal Oils Assoc.* 34, 125–129.
- Liu, D., Li, M., Ji, Y., Fu, Q., Li, M., Abrar Faiz, M., et al. (2021). Spatial-temporal characteristics analysis of water resource system resilience in irrigation areas based on a support vector machine model optimized by the modified gray wolf algorithm. *J. Hydrol.* 597:125758. doi: 10.1016/j.jhydrol.2020.125758
- Lu, W. (2006). *Modern Near Infrared Spectroscopy Analytical Technology*, 2nd Edn. Beijing: China Petrochemical Press.
- Lu, X., Adedze, Y. M. N., Chofong, G. N., Gandeka, M., Deng, Z., Teng, L., et al. (2018). Identification of high-efficiency SSR markers for assessing watermelon genetic purity. *J. Genet.* 97, 1295–1306. doi: 10.1007/s12041-018-1027-4
- Lun Liu, S., Sheng Tsai, Y., and Shau-mei, A. O. (2010). Classifying the variety, production area and season of Taiwan partially fermented tea by near infrared spectroscopy. *J. Food Drug Anal.* 18, 34–43.
- Luo, H., Zhu, J., Xu, W., and Cui, M. (2019). Identification of soybean varieties by terahertz spectroscopy and integrated learning method. *Optik* 184, 177–184. doi: 10.1016/j.ijleo.2019.02.148
- Ndlovu, P. F., Magwaza, L. S., Tesfay, S. Z., and Mphahlele, R. R. (2021). Rapid spectroscopic method for quantifying gluten concentration as a potential biomarker to test adulteration of green banana flour. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 262:120081. doi: 10.1016/j.saa.2021.120081
- Rezazad Bari, L., Ghanbari, A., Darvishzadeh, R., Giglou, M. T., and Baneh, H. D. (2021). Discernment of grape rootstocks base on their response to salt stress using selected characteristics in combination with chemometric tools. *Food Chem.* 365:130408. doi: 10.1016/j.foodchem.2021.130408
- Rong, D., Wang, H., Ying, Y., Zhang, Z., and Zhang, Y. (2020). Peach variety detection using VIS-NIR spectroscopy and deep learning. *Comput. Electron. Agric.* 175:105553. doi: 10.1016/j.compag.2020.105553
- Tafintseva, V., Shapaval, V., Blazhko, U., and Kohler, A. (2021). Correcting replicate variation in spectroscopic data by machine learning and model-based pre-processing. *Chemometr. Intell. Lab. Syst.* 215:104350. doi: 10.1016/j.chemolab.2021.104350
- Teye, E., Huang, X., Takrama, J., and Haiyang, G. (2014). Integrating NIR spectroscopy and electronic tongue together with chemometric analysis for accurate classification of cocoa bean varieties. *J. Food Process Eng.* 37, 560–566. doi: 10.1111/jfpe.12109
- Ujiie, A., Yamada, T., Fujimoto, K., Endo, Y., and Kitamura, K. (2005). Identification of soybean varieties with high α -tocopherol content. *Breed. Sci.* 55, 123–125. doi: 10.1270/jsbbs.55.123
- Wang, B., Gao, Y., Yuan, X., and Xiong, S. (2020a). Local R-symmetry Co-occurrence: characterising leaf image patterns for identifying cultivars. *IEEE ACM Trans. Comput. Biol. Bioinform.* 5963, 1–14. doi: 10.1109/tcbb.2020.3031280
- Wang, B., Gao, Y., Yuan, X., Xiong, S., and Feng, X. (2020b). From species to cultivar: soybean cultivar recognition using joint leaf image patterns by multiscale sliding chord matching. *Biosyst. Eng.* 194, 99–111. doi: 10.1016/j.biosystemseng.2020.03.019
- Wang, H., Xin, Y., and Wan, X. (2021a). Spectral detection technology of vegetable oil: spectral analysis of porphyrins and terpenoids. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 261:119965. doi: 10.1016/j.saa.2021.119965
- Wang, Y. T., Ren, H. B., Liang, W. Y., Jin, X., Yuan, Q., Liu, Z. R., et al. (2021b). A novel approach to temperature-dependent thermal processing authentication for milk by infrared spectroscopy coupled with machine learning. *J. Food Eng.* 311:110740. doi: 10.1016/j.jfoodeng.2021.110740
- Wang, R., Song, Q., Liu, Z., Ma, H., and Liu, Z. (2022). Multi-condition identification in milling Ti-6Al-4V thin-walled parts based on sensor fusion. *Mech. Syst. Signal Process.* 164:108264. doi: 10.1016/j.ymssp.2021.108264
- Wei, X., Li, S., Zhu, S., Zheng, W., Xie, Y., Zhou, S., et al. (2021a). Terahertz spectroscopy combined with data dimensionality reduction algorithms for quantitative analysis of protein content in soybeans. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 253:119571. doi: 10.1016/j.saa.2021.119571
- Wei, X., Li, S., Zhu, S., Zheng, W., Zhou, S., Wu, W., et al. (2021b). Quantitative analysis of soybean protein content by terahertz spectroscopy and chemometrics. *Chemometr. Intell. Lab. Syst.* 208:104199. doi: 10.1016/j.chemolab.2020.104199
- Wei, X., Zheng, W., Zhu, S., Zhou, S., Wu, W., and Xie, Z. (2020). Application of terahertz spectrum and interval partial least squares method in the identification of genetically modified soybeans. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 238:118453. doi: 10.1016/j.saa.2020.118453
- Wen, G., Dang, J., Xie, Z., Wang, J., Jiang, P., Guo, Q., et al. (2020). Molecular karyotypes of loquat (*Eriobotrya japonica*) aneuploids can be detected by using SSR markers combined with quantitative PCR irrespective of heterozygosity. *Plant Methods* 16:22. doi: 10.1186/s13007-020-00568-7

- Wu, G., Zhang, Y., Jiang, Y., Ge, H., and Lian, F. (2020). Identification of sesame oil varieties based on terahertz time domain spectroscopy. *Sci. Technol. Food Ind.* 41, 200–204. doi: 10.13386/j.issn1002-0306.2020.04.034
- Xue, S. S., Tan, J., Xie, J. Y., and Li, M. F. (2021). Rapid, simultaneous and non-destructive determination of maize flour and soybean flour adulterated in quinoa flour by front-face synchronous fluorescence spectroscopy. *Food Control* 130:108329. doi: 10.1016/j.foodcont.2021.108329
- Yang, S., Li, C., Mei, Y., Liu, W., Liu, R., Chen, W., et al. (2021). Discrimination of corn variety using terahertz spectroscopy combined with chemometrics methods. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* 252:119475. doi: 10.1016/j.saa.2021.119475
- Yang, X., Zhao, X., Yang, K., Liu, Y., Liu, Y., Fu, W., et al. (2016). Biomedical applications of terahertz spectroscopy and imaging. *Trends Biotechnol.* 34, 810–824. doi: 10.1016/j.tibtech.2016.04.008
- Zhou, J., Huang, S., Wang, M., and Qiu, Y. (2021a). Performance evaluation of hybrid GA-SVM and GWO-SVM models to predict earthquake-induced liquefaction potential of soil: a multi-dataset investigation. *Eng. Comput.* doi: 10.1007/s00366-021-01418-3
- Zhou, W., Chen, M., Yang, Z., and Song, X. (2021b). Real estate risk measurement and early warning based on PSO-SVM. *Socioecon. Plan Sci.* 77:101001. doi: 10.1016/j.seps.2020.101001
- Zhu, D., Wang, K., Zhou, G., Hou, R., and Wang, C. (2010). The NIR spectra based variety discrimination for single soybean seed. *Spectrosc. Spectr. Anal.* 30, 3217–3221. doi: 10.3964/j.issn.1000-0593201012-3217-05
- 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 Wei, Kong, Zhu, Li, Zhou and Wu. 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 Prediction of Green Fraction Dynamics in Soybean Using Unmanned Aerial Vehicles Observations

Yusuke Toda¹, Goshi Sasaki¹, Yoshihiro Ohmori¹, Yuji Yamasaki², Hirokazu Takahashi³, Hideki Takanashi¹, Mai Tsuda⁴, Hiromi Kajiya-Kanegae⁵, Raul Lopez-Lozano⁶, Hisashi Tsujimoto², Akito Kaga⁷, Mikio Nakazono³, Toru Fujiwara¹, Frederic Baret⁶ and Hiroyoshi Iwata^{1*}

¹ Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan, ² Arid Land Research Center, Tottori University, Tottori, Japan, ³ Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya, Japan, ⁴ Tsukuba-Plant Innovation Research Center (T-PIRC), University of Tsukuba, Tsukuba, Japan, ⁵ Research Center for Agricultural Information Technology, National Agriculture and Food Research Organization (NARO), Tokyo, Japan, ⁶ Joint Research Unit of Mediterranean Environment and Modelling of Agroecosystems, National Research Institute for Agriculture, Food and Environment (INRAE), Avignon, France, ⁷ Institute of Crop Science, National Agriculture and Food Research Organization (NARO), Tsukuba, Japan

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Hilde Muylle,
Institute for Agricultural, Fisheries
and Food Research (ILVO), Belgium
Zitong Li,
Commonwealth Scientific
and Industrial Research Organisation
(CSIRO), Australia

*Correspondence:

Hiroyoshi Iwata
aiwata@mail.ecc.u-tokyo.ac.jp

Specialty section:

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

Received: 10 December 2021

Accepted: 21 February 2022

Published: 16 March 2022

Citation:

Toda Y, Sasaki G, Ohmori Y,
Yamasaki Y, Takahashi H,
Takanashi H, Tsuda M,
Kajiya-Kanegae H, Lopez-Lozano R,
Tsujimoto H, Kaga A, Nakazono M,
Fujiwara T, Baret F and Iwata H (2022)
Genomic Prediction of Green Fraction
Dynamics in Soybean Using
Unmanned Aerial Vehicles
Observations.
Front. Plant Sci. 13:828864.
doi: 10.3389/fpls.2022.828864

With the widespread use of high-throughput phenotyping systems, growth process data are expected to become more easily available. By applying genomic prediction to growth data, it will be possible to predict the growth of untested genotypes. Predicting the growth process will be useful for crop breeding, as variability in the growth process has a significant impact on the management of plant cultivation. However, the integration of growth modeling and genomic prediction has yet to be studied in depth. In this study, we implemented new prediction models to propose a novel growth prediction scheme. Phenotype data of 198 soybean germplasm genotypes were acquired for 3 years in experimental fields in Tottori, Japan. The longitudinal changes in the green fractions were measured using UAV remote sensing. Then, a dynamic model was fitted to the green fraction to extract the dynamic characteristics of the green fraction as five parameters. Using the estimated growth parameters, we developed models for genomic prediction of the growth process and tested whether the inclusion of the dynamic model contributed to better prediction of growth. Our proposed models consist of two steps: first, predicting the parameters of the dynamics model with genomic prediction, and then substituting the predicted values for the parameters of the dynamics model. By evaluating the heritability of the growth parameters, the dynamic model was able to effectively extract genetic diversity in the growth characteristics of the green fraction. In addition, the proposed prediction model showed higher prediction accuracy than conventional genomic prediction models, especially when the future growth of the test population is a prediction target given the observed values in the first half of growth as training data. This indicates that our model was able to successfully combine information from the early growth period with phenotypic data from the training population for prediction. This prediction method could be applied to selection at an early growth stage in crop breeding, and could reduce the cost and time of field trials.

Keywords: soybean, unmanned aerial vehicle, remote sensing, drought, green fraction, dynamics model, genomic prediction, genomic selection

INTRODUCTION

Genetic mechanisms of growth processes have become a crucial topic in plant breeding. The genetic dissection of the formation process of target traits of breeding such as yield quantity and quality will provide profound insights into its mechanism, which will lead to efficient selection of useful genotypes and rapid genetic improvement. This understanding is important in genomic selection (GS), where breeders skip field trials and select promising candidates based on the predicted breeding value provided by genomic prediction (GP) (Meuwissen et al., 2001). Most GP studies on crops have focused on traits at harvest, such as yield and quality (Krishnappa et al., 2021). If GP can predict genetic variation during the growth process, breeders can accurately determine the behavior of the genotypes obtained by GP and select the appropriate candidates. Further, growth prediction from the early period is likely to reduce the cost of field trials by shortening the period of observation.

However, the measurements required to derive data of growth trajectory are cost and labor intensive, representing a major bottleneck for genetic dissection, which requires the characterization of many genotypes. Due to the rapid development of sensing technologies in recent years, high-throughput phenotyping has become available for plant breeding, and the measurement of growth traits is becoming more practical. An accurate and detailed acquisition of growth processes through high-throughput measurements is expected to lead to improved genetic gains in plant breeding (Furbank and Tester, 2011; Cabrera-Bosquet et al., 2012; Araus and Cairns, 2014). For example, an automated phenotyping platform for the monitoring of three-dimensional plant growth in a greenhouse has enabled the genetic dissection of growth processes using a dynamic model (Campbell et al., 2018). In a field experiment, high-throughput phenotyping using unmanned aerial vehicles (UAVs) (Yang et al., 2017) and tractors (White et al., 2012) was used to measure plant growth. Among growth traits, the leaf area index (LAI) is often investigated because it is accessible from high-throughput phenotyping (Verger et al., 2014; Liu et al., 2017) while being sensitive to the environment, directly determining amount of light absorption, and thus affecting biomass production and yield. Until recently, however, these techniques were mainly used for crop management such as estimation of canopy state variables, soil properties and yield (Jin et al., 2018), and their applications to genetic dissection remain limited (Blancon et al., 2019).

Several methods have been proposed for the analysis of plant growth. One commonly used method involves fitting a growth model, such as Gompertz (Winsor, 1932) and logistic (Nelder, 1961), to the data and using the model parameters to quantify the dynamic pattern. This method can be applied to various types of dynamic measurements such as stem diameter of trees (Wu et al., 2004) and soybean canopy cover and height (Borra-Serrano et al., 2020). Several methods of quantitative genetics, such as quantitative trait loci analysis (Ma et al., 2002; Wu et al., 2002) and genome-wide association studies (Das et al., 2011; Crispin et al., 2015), have been applied to discover possible associations with growth model parameters.

Growth models have also been used as a flexible tool to analyze various factors, such as the effect of selection in breeding (Piles et al., 2003) and the relationship among traits (Onogi et al., 2019). However, its application to GP has not been discussed in previous studies.

In this study, a method integrating a model of growth dynamics and GP was proposed and applied to investigate the growth of soybeans. We focused on the green fraction (GF) to model its dynamics. GF is defined as the fraction of green pixels in an image taken from the sky. This trait is a proxy for LAI and can be easily measured from UAV observations. The GF dynamics of soybean germplasm accessions were described using the parameters of a model consisting of logistic and exponential curves. Genetic variations in GF dynamics were quantified by decomposing the model parameters into genetic and residual effects using mixed models. Finally, the GP model is applied to predict the parameters of the GF dynamics model under a range of scenarios to illustrate the potential of the proposed method. A similar experiment was conducted in an earlier paper in which UAV-RS data was used as secondary traits to predict biomass (Toda et al., 2021a), while this study developed prediction models of growth curve itself.

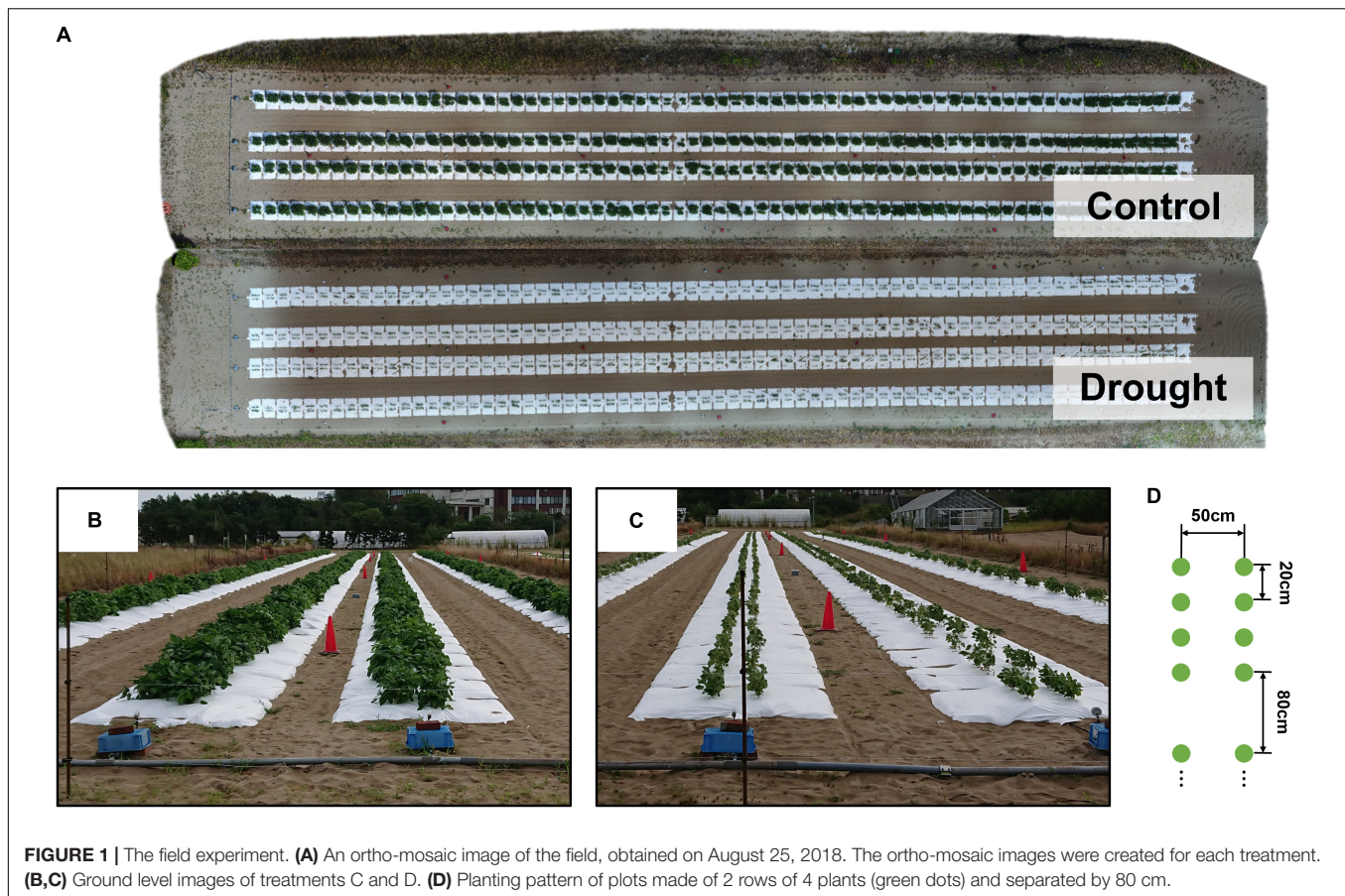
MATERIALS AND METHODS

Field Trials

Soybean accessions registered in the National Agriculture and Food Research Organization Genebank¹ were used. A total of 198 accessions, consisted of 96 Japanese accessions and 96 world accessions from mini core collection (Kaga et al., 2012) and 6 additional accessions. From 2017 to 2019, the field trial was conducted in an experimental field with sandy soil at the Arid Land Research Center, Tottori University (35°32' N lat, 134°12' E long, 14 m above sea level) (**Supplementary Figure 1**). A total of 198 accessions between 2018 and 2019 were used, with 186 out of 198 accessions used in 2017. Each plot consisted of four plants. The distances between two rows, two plots, and two individuals were 50, 80, and 20 cm, respectively (**Figure 1D**). Sowing was performed at the beginning of July, followed by thinning after 2 weeks (**Supplementary Table 1**). Fertilizer (15, 6.0, 20, 11, and 7.0 g m⁻² of N, P, K, Mg, and Ca, respectively) was applied to the field before sowing.

Two watering treatment levels, control (C) and drought (D), were used to evaluate the genetic variations in the responses to water stress. White mulching sheets (Tyvek, Dupond, United States) were laid to prevent rainwater infiltration (**Figure 1**), and pipes were installed under the sheets to irrigate the field. Irrigation with irrigation rate 8.1 mm/h was applied for daily for 5 h (7:00–9:00, 12:00–14:00, 16:00–17:00), starting the day after the thinning in treatment C, while no water was brought in treatment D. In the following, an abbreviation for denoting a specific combination of the level of the treatment and the year of the experiment is used; for example, treatment C in 2017 is abbreviated to “2017-C.”

¹https://www.gene.affrc.go.jp/index_en.php



Remote Sensing and Image Analysis

UAV flights started after thinning and were performed 16–35 times during the cultivation period. A consumer drone (DJI Phantom 4 Advanced, China) was used for image collection. Images consisted of RGB layers and $3,648 \times 4,864$ pixels, captured with an automated focus and white balance. The UAV flew 12–14 m above the ground and captured images every 2 s with an autofocus function. A single UAV flight took approximately 15 min and collected 500–600 images, which was repeated twice to cover the entire field.

Ortho-mosaic images were constructed using Pix4Dmapper (Pix4D, Switzerland). The images of individual plots were then segmented from the ortho-mosaic image based on the geolocation of their corners. The canopy regions of the images of the individual plots were segmented based on GRVI and hue values ($\text{GRVI} < 0.05$, $20 < \text{Hue} < 90$). Finally, the GF of each plot was estimated as the ratio of the green pixels to the total number of pixels in the plot. The image analysis process was implemented in Python 3.7² and library opencv (ver.4.1.0) and gdal (ver.3.2.2). For data in 2019, a similar procedure was used by Hiphen Inc.³ The analysis protocol was the same as previous research (Verger et al., 2014; Madec et al., 2017).

²<https://www.python.org>

³<https://www.hiphen-plant.com>

Green Fraction Dynamics Modeling

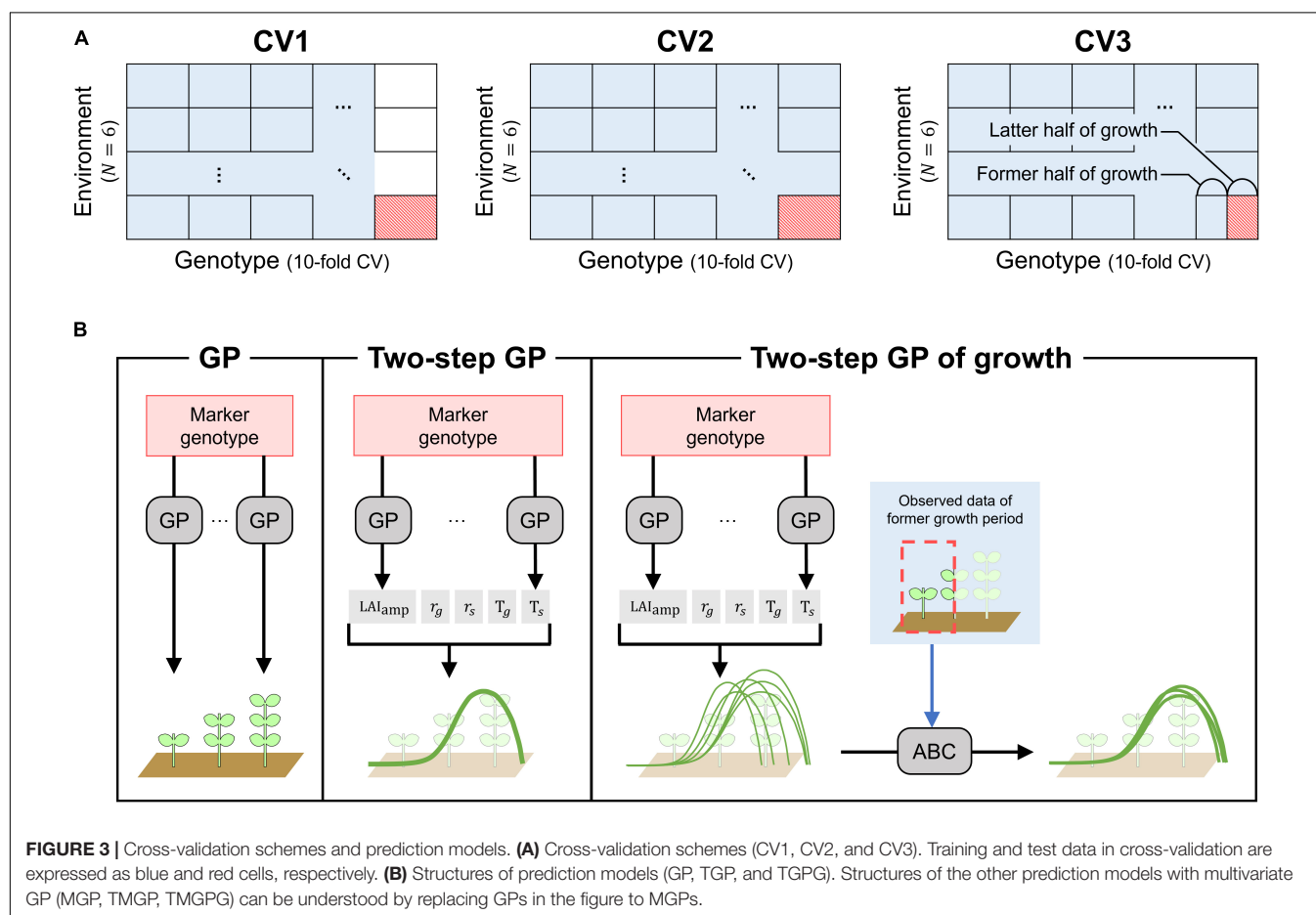
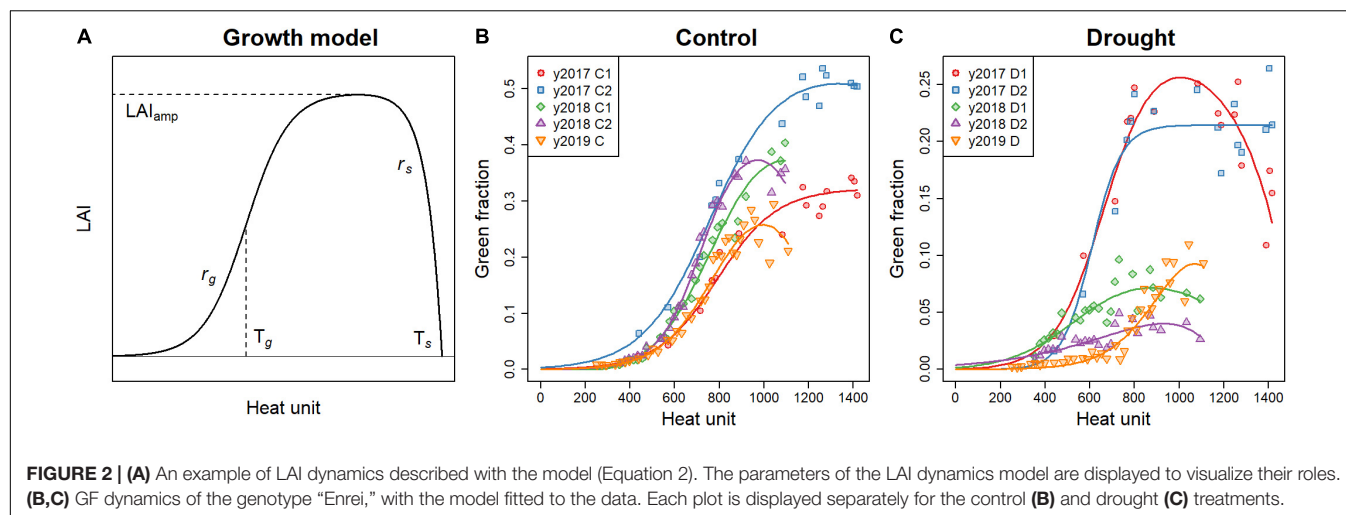
The GF derived from the UAV on day d day, GF_d , was first converted into the corresponding leaf area index (LAI_d), following (Soltani and Sinclair, 2012) the exponential model:

$$\text{GF}_d = 1 - \exp(-k \text{LAI}_d) \quad (1)$$

where $k = 0.5$, and is the extinction coefficient commonly used for soybean. The model proposed by Koetz et al. (2005) to describe the dynamics of LAI was fitted to the time series of GF to estimate the growth pattern of each plot with five parameters (**Figure 2**):

$$\text{LAI}_d = \text{LAI}_{\text{amp}} \left\{ \frac{1}{1 + \exp(-r_g (T_d - T_g))} - \exp(r_s (T_d - T_s)) \right\} \quad (2)$$

The first term in the parenthesis represents logistic growth, and the second term represents exponential senescence, where T_d is a growing degree day on day d . The growing degree day is a typical development scale corresponding to the cumulative daily mean temperature from sowing above the base temperature set to 8°C (Soltani and Sinclair, 2012). LAI_{amp} is the maximum value of the LAI reached, r_g is the maximum LAI growth rate, r_s is the senescence rate, T_g is the growing degree day when the LAI growth rate is maximum, and T_s is the growing degree day when LAI becomes zero. The five parameters, LAI_{amp} , r_g , r_s , T_g , and T_s , were estimated for each plot. However,



fitting the dynamics model separately for each plot was difficult because the GF in the drought treatment (D) was so small. Many of the GF growth data from the treatment D contained large noise, making it difficult to estimate the parameters, from data of each plot alone, especially the inflection point of the growth T_g . Therefore, the estimation was conducted in two steps:

(1) First parameter estimates.

In this step, all of the parameters except LAI_{amp} were assumed to be genotype dependent and treatment independent, that is, parameters of each plot were the same for treatments C and D if their genotypes were the same. The optimal values of T_g and T_s were found with a grid search in the range $300 < T_g < 1,200$ and $1,400 < T_s < 3,000$ on seven points evenly distributed in

the range. At the same time, the optimal values of the other parameters were estimated using the Nelder-Mead method. The cost function to be minimized in the Nelder-Mead method was computed as follows:

$$\sum_d \sum_i \frac{(y_{i,d} - \hat{y}_{i,d})^2}{\bar{y}_d}, \quad (3)$$

where $y_{i,d}$ is the GF of plot i on day d , $\hat{y}_{i,d}$ is the estimated value of GF with the dynamics model, and \bar{y}_d is the mean value of the GF on day d . The normalization by \bar{y}_d in Equation 3 accounts for the measurement noise, which is roughly proportional to the mean value.

(2) Fine tuning the parameter estimates.

The parameter estimation was conducted independently for each plot. The optimal values obtained in the previous step for T_g and T_s were the center of the grid search with a narrower range ($T_g - 200 < T_g < T_g + 200$ and $T_s - 400 < T_s < T_s + 400$). The other parameters were estimated using the Nelder-Mead method, using the estimated values in Step 1 as the initial values.

Estimation of Genotypic Values

Genotypic values of the GF and LAI dynamic parameters were estimated for use in the GP. The following mixed model was fitted for each combination of a trait (GF or LAI dynamics parameter) and a treatment (C or D):

$$\mathbf{y} = \mu \mathbf{1} + \mathbf{L}\boldsymbol{\beta} + \mathbf{W}\mathbf{s} + \mathbf{e} \quad (4)$$

where \mathbf{y} is a vector of the phenotypic values; μ is the mean; $\boldsymbol{\beta}$ is a vector of block effects representing differences between replications; \mathbf{s} is a vector of genotypic values that follows $N(\mathbf{s} | \mathbf{0}, \sigma_s^2 \mathbf{I})$; σ_s^2 is the genotypic variance; \mathbf{e} is a vector of residuals that follows $N(\mathbf{e} | \mathbf{0}, \sigma_e^2 \mathbf{I})$; σ_e^2 is the residual variance; $\mathbf{1}$ is a vector in which all the elements are one; \mathbf{I} is an identity matrix; and \mathbf{L} and \mathbf{W} are design matrices. The genotypic value (\mathbf{g}) was calculated as follows:

$$\mathbf{g} = \mu \mathbf{1} + \mathbf{s} \quad (5)$$

The R package lme4 (ver. 1.1–20) was used to solve Equation 4. For the GF, the genotypic value estimation was applied separately for each flight date.

Genomic Relationship Matrix and Genetic Analysis

The whole-genome sequencing data of all 198 accessions were available and used to estimate the genomic relationship matrix (Kajiya-Kanegae et al., 2021). Only the biallelic sites in all accessions with a minor allele frequency (MAF) ≥ 0.025 , missing rate < 0.05 , and linkage disequilibrium < 0.95 were extracted, and the imputation of missing genotypes was applied. Genotyping data identified 425,858 SNPs. Genotypes for individual alleles were represented as -1 (homozygous for the reference allele), 1 (homozygous for the alternative allele), or 0 (heterozygous for the reference and alternative alleles). The genomic relationship matrix \mathbf{G} was estimated as $\mathbf{G} = \mathbf{X}\mathbf{X}^T / c$, where \mathbf{X} is an $n \times m$ scaled marker genotype matrix (n and m

are the numbers of lines and markers, respectively), and c is the normalization constant (Endelman and Jannink, 2012). Genetic heritability was estimated for all traits using the genomic best linear unbiased prediction (G-BLUP) model:

$$\mathbf{g} = m\mathbf{1} + \mathbf{Z}\mathbf{u} + \mathbf{e} \quad (6)$$

where \mathbf{g} is a vector of genotypic values estimated using Equations 4 and 5, m is the mean, \mathbf{u} is a vector of random genetic effects that follows $N(\mathbf{u} | \mathbf{0}, \sigma_u^2 \mathbf{G})$, \mathbf{e} is a vector of residuals that follows $N(\mathbf{e} | \mathbf{0}, \sigma_e^2 \mathbf{I})$, σ_u^2 and σ_e^2 are the genetic and residual variances, respectively, and \mathbf{Z} is a design matrix. The R package rrBLUP (ver. 4.6) (Endelman, 2011) was used to solve Equation 6. After solving the mixed model, the genomic heritability was estimated as $h^2 = \sigma_u^2 / (\sigma_u^2 + \sigma_e^2)$.

Prediction of Green Fraction Dynamics

We investigated three cross-validation schemes for the four different prediction models. The cross-validation schemes and prediction models are detailed as follows: The correlation coefficient between the genotypic values (\mathbf{g}) and their predicted values (\mathbf{u}) of the GF was used to evaluate the prediction accuracy.

Cross-Validation Schemes

Cross-validation was repeated three times for the combination of a cross-validation scheme and a prediction model.

(1) *Cross-validation of genotypes (CV1).*

CV1 corresponded to the prediction of LAI dynamics for untested genotypes. Data from a subset of genotypes in any treatment or year were excluded from the training data (Figure 3A). The prediction model built using the training dataset was evaluated for the left-out genotypes. Ten-fold cross-validation was used to randomly select 19–20 left-out genotypes.

(2) *Cross-Validation over combination of genotype and environment (CV2).*

The combination of a treatment and a year was considered as an environment: there were a total of six environments (two treatments \times 3 years). Here, the predicted LAI dynamics were evaluated for genotypes and environments left out from the training dataset. A 10-fold cross-validation of genotypes and leave-one-environment-out cross-validation were applied simultaneously to get rid of data of test genotypes in one environment from training dataset (Figure 3A).

(3) *Cross-validation with a focus on the late growth period (CV3).*

The growth cycle was split into early and late growth, with an equal number of observations for the two periods. CV3 was similar to CV2, but data of early growth period of test genotypes in a test environment was included in training data, and prediction of the LAI dynamics were evaluated over the late period (Figure 3A).

Prediction Models in Cross-Validation of Genotypes and Cross-Validation Over Combination of Genotype and Environment

In CV1 and CV2, four prediction models were compared: genomic prediction (GP), two-step GP (TGP), multivariate GP (MGP), and two-step multivariate GP (TMGP).

GP is the most standard model, expressed as shown in Equation 6, and applied to the GF on each day in the training data (**Figure 3B**). Then, the random genetic values g of the left-out genotypes were used as the predicted values.

The TGP consisted of two steps (**Figure 3B**). First, the same model as GP was applied to the LAI dynamics model parameters. Then, the GFs of the left-out genotypes on each day were calculated using the predicted parameters (Equations 1 and 2).

MGP is an extension of the GP, which simultaneously predicts several traits (Calus and Veerkamp, 2011; Jia and Jannink, 2012). The model is expected to enhance the accuracy of genomic prediction *via* genetic correlations among traits. This model can be expressed as follows:

$$\begin{pmatrix} \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_J \end{pmatrix} = \begin{pmatrix} \mathbf{m}_1 \mathbf{1} \\ \vdots \\ \mathbf{m}_J \mathbf{1} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_1 & \cdots & \mathbf{O} \\ \vdots & \ddots & \vdots \\ \mathbf{O} & \cdots & \mathbf{Z}_J \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_J \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_J \end{pmatrix} \quad (7)$$

where J is the number of variates in the model, \mathbf{g}_j , \mathbf{u}_j , and $\boldsymbol{\varepsilon}_j$ are vectors of genotypic values, random genetic effects, and residuals of variate j , respectively, and m_j is the mean of variate j . Assumptions for the random effects were included, in which $\mathbf{u}_{\text{all}} = (\mathbf{u}_1^T, \dots, \mathbf{u}_J^T)^T$ follows $N(\mathbf{u}_{\text{all}} | \mathbf{0}, \mathbf{K} \otimes \mathbf{G})$ and $\boldsymbol{\varepsilon}_{\text{all}} = (\boldsymbol{\varepsilon}_1^T, \dots, \boldsymbol{\varepsilon}_J^T)^T$ follows $N(\boldsymbol{\varepsilon}_{\text{all}} | \mathbf{0}, \mathbf{R} \otimes \mathbf{I})$. Here, \mathbf{K} is a genomic variance-covariance matrix between the variates, and \mathbf{R} is the residual variance-covariance matrix between the variates. The R package MTM (ver. 1.0.0) was used to solve Equation 7 based on the Markov chain Monte Carlo (MCMC) method.

TMGP consisted of two steps, that is, MGP of the LAI dynamics model parameters and the calculation of the GF using the predicted parameters.

MGP and TMGP were expected to improve the prediction accuracy compared to GP by exploiting phenotypic data from environments included only in the training dataset. However, since the GF was measured repeatedly in each environment, it was difficult to include all the phenotype data (152 measurements in total by adding up observation dates in all the environments). Thus, a strategy was applied where the training of prediction models was repeated for each observation date, and ten additional variates were selected from the whole data to support the prediction every time. In other words, the eleven variates included each time consisted of one target variate and ten supporting variates. The criterion for selecting supporting variates is based on heritability and correlation with the target variate. These two factors are essential for improving the prediction accuracy in MGP (Calus and Veerkamp, 2011). Top-10 observations of the following criterion were selected as supporting variables:

$$s(h^2) + s(|r|) \quad (8)$$

where $s(\cdot)$ is a scaling function that makes the mean and variance of an input vector zero and one, respectively; h^2 is the heritability; and r is the correlation coefficient with a target variate.

The Prediction Models in Cross-Validation With a Focus on the Late Growth Period

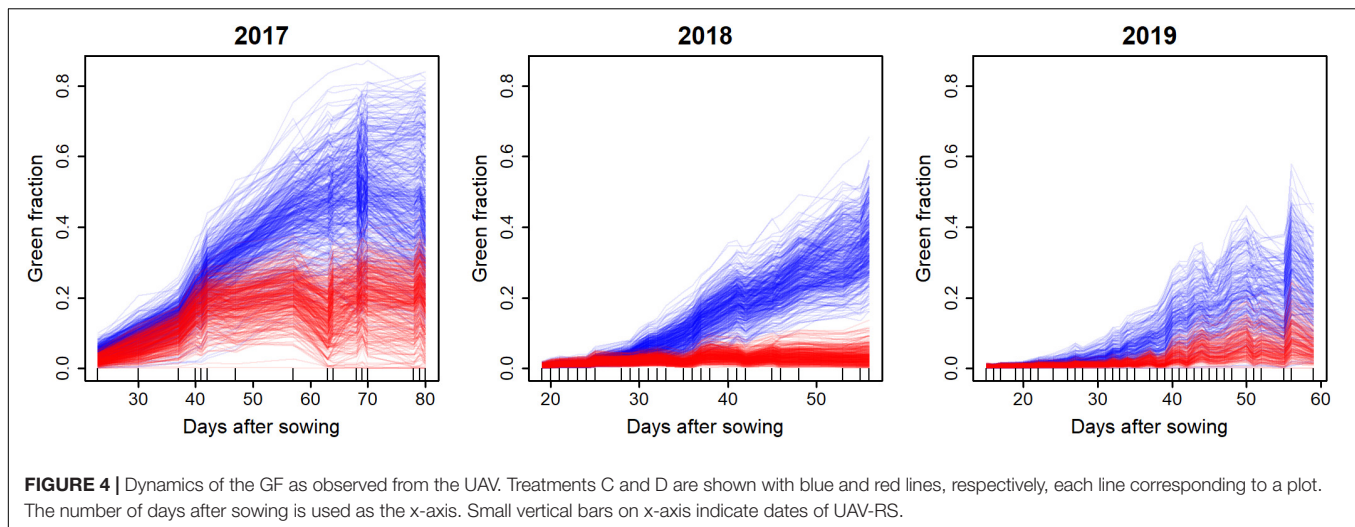
As in the other cross-validations, the performances of the four prediction models were compared in CV3. GP was the same as in CV2 because it only uses the data of the measurement day to be predicted for training. MGP was modified to better exploit the first half of the growth period used to train the model. Seven out of the ten supporting variates were selected using the selection criterion from Equation 8, the remaining three variates corresponded to the GF values for the latest three flights of the first growth period. The other two models with two-step structures, TGP and TMGP, were also modified to better exploit the training data for predicting the GF dynamics during the late growth period. They were called TGPG (TGP for growth) and TMGPG (TMGP for growth), respectively.

The TGPG included three steps (**Figure 3B**). The first two steps were the same as those of the TGP, where the LAI dynamics model parameters were predicted without using the data from the first half of the growth period. However, for TGPG, the distributions of the MCMC values of the LAI dynamics model parameters were used instead of the average value of the samples used in the TGP. As a result, 60,000 samples of the predicted GF dynamics were obtained for each genotype corresponding to the prior distribution when no GF measurements on the genotype were available. Then, the GF data from the first half of the growth period were exploited using the approximate Bayesian computation (ABC) method, and the 60 GF dynamics that minimize the Euclidian distance between the predicted GF dynamics and the actual GF observations were selected. Lastly, the mean values of the 60 samples were used as the predicted values. The modifications on TMGP to obtain TMGPG were the same as those applied to TGP to obtain TGPG.

RESULTS

Dynamics of Green Fraction Derived From Unmanned Aerial Vehicles

The UAV observations transformed into GF values show typical dynamics (growth, saturation, and senescence) of the several genotypes and environments investigated (**Figure 4**), showing large variations in the growth patterns. It is worth noting that the period covered by the flights was longer in 2017, with up to 80 days compared to 2018 and 2019, where the flights were stopped after 60 days. For each plot, the GF dynamics were relatively smooth, indicating a good temporal consistency of the GF values derived from the UAV observations. The ranking between genotypes is also generally consistent across growth development, which would indicate good chances to predict the late period from observations covering the early growth period. Drought treatment (D) always showed lower GF values than the control (C) treatment. However, the water stress experienced by treatment D varied across years, with 2018 being the most severe, and 2017 the mildest. The control treatment also showed differences between the 3 years: 2017 showed the best growth conditions, while 2019 showed the worst ones.



The genomic heritability estimated for each year and treatment by fitting a mixed model to GF for each observation day showed a decrease until 40–50 days after sowing, and then increased with time (**Figures 5A,B**). However, some differences between years were observed, with a higher heritability in the early stages of 2018. The yearly patterns were also similar between the control (C) and drought (D) treatments, while the heritability in treatment C was systematically higher than that in treatment D, except in 2017 for the late UAV flights.

Growth Parameter Estimation

The dynamics model fitted a wide range of GF growth patterns in both treatments and all years (**Figure 2**). The root mean squared errors (RMSE) of growth model fitting of GF on 25 days after sowing were 0.0060 and 0.0057 in treatment C and D, respectively. RMSE reached 0.022 and 0.016 in two treatments on 50 days after sowing, because growth of canopy increased the measurement noise of GF.

The distribution of the estimated growth parameters varied among years and treatments (**Figure 6**). Fundamentally, the parameters related to period of growth (r_g , T_g , and LAI_{amp}) showed a tendency wherein the values of the parameters became smaller when the plants were subjected to drought stress. For the parameters related to period of senescence (r_s , T_s), the results were not reliable due to the lack of observation of senescence, except in 2017.

The genomic heritability of the growth parameters (**Figures 5C,D**) varied among treatments and years. The heritabilities of LAI_{amp} , T_g , and r_g were relatively high, reaching 0.8 in the highest cases. The other two parameters, r_s and T_s , showed lower heritability, ranging between 0.1 and 0.5. These parameters characterize the late development of the canopy that was not well covered by UAV flights, except in 2017. This explains why heritability was highest in 2017. Particularly, heritability in 2017-D exceeded 0.4 with all values. On the other hand, heritability of all the parameters was lower than 0.4 in 2018-D when summer heat stress was severe.

The heritability of the parameters of the model is generally lower than that of the GF for each UAV flight (**Figure 5**), except for LAI_{amp} , which generally shows a higher heritability, except in 2018.

Prediction of Growth Patterns

In CV1, the prediction accuracy of TGP and MGP was similar to that of GP (**Figure 7**). In 2019-D, a significant improvement in prediction accuracy in MGP (50.0% improvement in ratio of correlation coefficients of genotypic and predicted values) compared with GP was observed, where the accuracy of GP was very low in the latter half of the growth period (**Supplementary Figure 2**). The accuracy of MGP was higher (12.6%) than that of TGP. The accuracy of TMGP differed among environments; it was lower than that of GP when predicting the GF in 2018, while it was higher than the accuracy of MGP when predicting the GF in the latter half of the growth period in 2019.

The predicted values of GP and TGP in CV2 were equal to those in CV1 because they did not utilize data in environments other than their targets. Thus, for CV2, the focus will be on the accuracy of the MGP and TMGP. The accuracy of MGP was higher in CV2 than in CV1 (11.6%) and was significantly higher than that of TGP (25.6%) (**Figure 7** and **Supplementary Figure 3**). The accuracy of TMGP was lower in CV2 than in CV1 in 2018, while it was higher in CV2 than in CV1 in the other years. Comparing MGP and TMGP, the accuracy of MGP was higher in 2018 and the former half of the growth period, while that of TMGP was higher in other environments.

In CV3, the prediction accuracy of TGPG and TMGPG was higher than that of the other models (24.5 and 27.1%, compared with MGP) in all the environments (**Figure 7**) over the entire growth period (**Supplementary Figure 4**). The correlation coefficients between the predicted values of TMGPG and the genotypic values were higher than 0.6 in most cases. The accuracy of MGP was higher than that of GP (71.7%), but lower than that of TGPG and TMGPG. As in CV2, the predicted values with GP and TGP in CV3 were the same as in CV1.

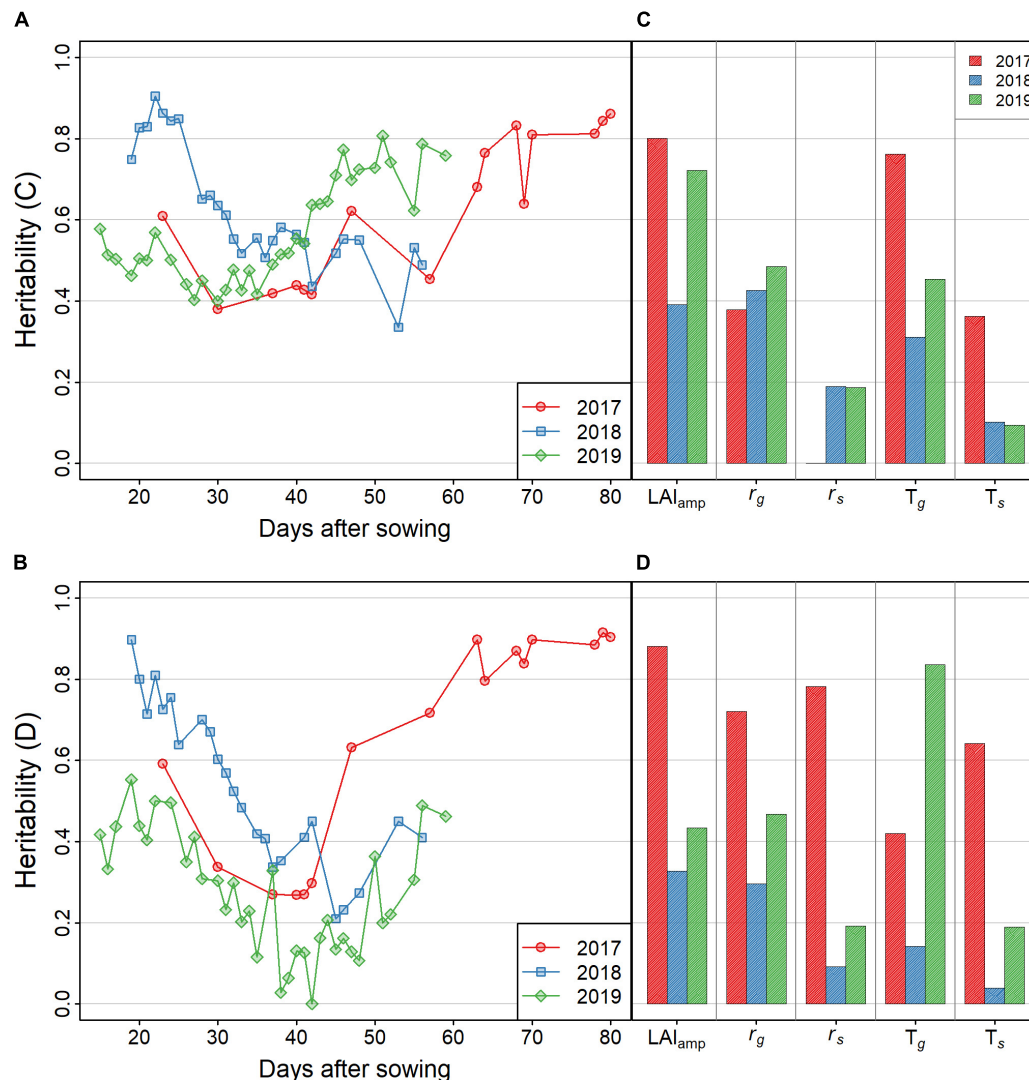


FIGURE 5 | (A,B) Heritability of the GF. Results of treatments C and D are shown in **(A,B)**, respectively. Red, blue, and green lines indicate the values in 2017, 2018, and 2019, respectively. **(C,D)** Heritability of the LAI dynamics model parameters. Results of treatments C and D are shown in **(C,D)**, respectively.

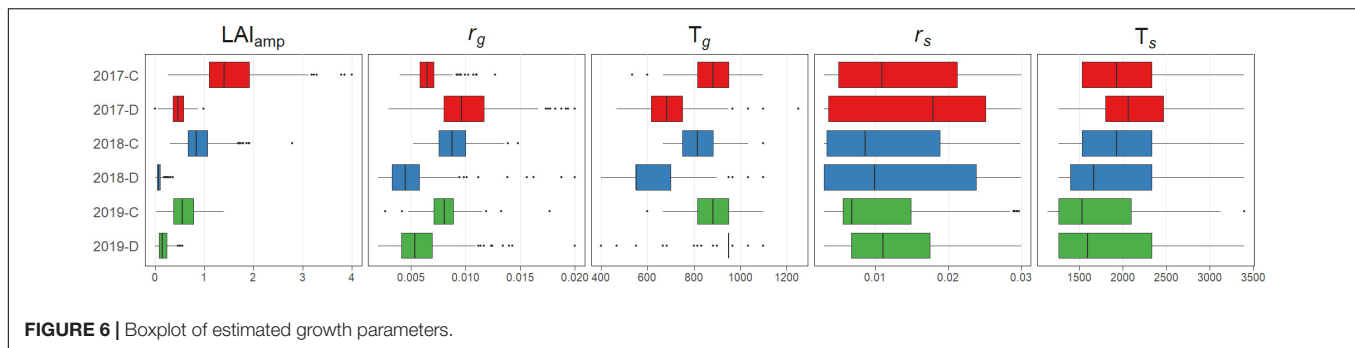
DISCUSSION

Unmanned Aerial Vehicles-RS as a Tool to Evaluate Growth Patterns

This study showed that the UAV measurements of GF could be used to assess the genetic diversity of soybean growth patterns. A U-shaped longitudinal pattern was observed in the heritability of GF in all environments (**Figures 5A,B**). The U-shaped heritability patterns can be explained in three steps. In the early stage of growth, the GF seemed to be determined by factors regarding initial growth speed, such as radiation use efficiency, which results in high heritability of the GF. At approximately 25 days after sowing, several additional factors, such as growth phenology and plant structure, related to phenological development started to affect the GF, which decreased the heritability. Then, the saturation of the GF

occurred around 45–60 days after sowing. During this period, confounding factors related to the differences in phenological development weakened, leading to increased heritability.

When we evaluate crop dynamics using remote sensing techniques, the LAI is often used as the main trait of interest. To apply remote sensing techniques to breeding populations, differences in plant styles should be carefully considered when deriving LAI from the images (Blancón et al., 2019). In this study, we focused on GF rather than LAI as the target trait to model its dynamics, due to the lack of data on plant style. Applying our method to LAI would offer the advantage of being closer to crop growth mechanisms, such as photosynthesis. Nevertheless, this study has shown that GF is a useful trait to describe the growth of soybean germplasm. In genomic prediction, the inclusion of canopy area, which is proportional to GF, has been reported to improve the prediction of biomass in soybean (Toda et al., 2021a).



Thus, GF can be considered a useful index of genetic variation in plant growth.

Fitting a Dynamics Model

The dynamics model was flexible enough to represent various growth patterns in different environments (Figure 2). By fitting the model, the GF time series is represented by five parameters. The distribution of growth parameters reflected the effects of drought stress on growth (Figure 6). The decrease in r_g and T_g in treatment D indicates that the speed and duration of growth were strongly suppressed under drought stress. As a result, the maximum value of the plant canopy was significantly reduced, which was expressed in the reduction of LAI_{amp} .

Genetic analysis of the parameters showed that the heritability of LAI_{amp} was the highest (Figures 5C,D). This result is related to the high heritability of GF in the later stages of growth. T_g , which describes the stage-shift timing of the GF, also showed high heritability in 2017-C and 2019-D.

For the senescence stage, the heritability was low for T_s and r_s , except for 2017-D. Due to the long cultivation period in 2017 and early senescence in treatment D, model fitting of the senescence part was successful in that environment. The heritability of r_s was higher than T_s , which means that the change in GF in the senescence stage was mainly determined by its speed, r_s , rather than the timing of senescence, T_s . The senescence pattern could be evaluated more precisely in 2017 and 2018 by extending the observation period.

Several useful results were obtained by applying the dynamics model, and some problems were found to be improved. Because of the high heritability of GF in the early stages, the growth speed, r_g , was expected to be mainly determined by genetic factors. However, the heritability of r_g was low, except in 2017-D. The use of other dynamic models, such as the Gompertz (Winsor, 1932) curve, may improve the goodness of fit of a growth curve to the GF in the early growth stages. It was reported that the dynamics model that considers leaf appearance could explain the dynamics of green LAI (GLAI) (Blanc et al., 2019). Such structural models may also be candidates for alternative dynamic models.

Another possible improvement of the model is the inclusion of other environmental factors, such as soil moisture and drought stress. In the dynamics model, the effect of temperature on growth stages was considered. However, the inclusion of other factors may allow for improved fitting and simultaneous

parameter estimation of multiple environments. For example, the low heritability in 2018 of the growth parameters was due to severe heat stress in the summer of 2018, which made the growth slower than usual years. As a result, the sigmoid pattern in growth was truncated at the end of the cultivation period (Figure 4). Other environmental factors will allow simultaneous parameter estimation in other environments, leading to stability in the estimated parameters.

Prediction of Growth Curves

In CV1, the accuracy was close between GP and TGP (Figure 7). This result suggests that the dynamics model used in TGP could extract sufficient genetic variations from phenotypic variations in the GF dynamics pattern to achieve the same predictive accuracy as the GP.

Models with multivariate GP yielded better accuracy than those with univariate GP; the accuracies of MGP and TMGP were higher than those of GP and TGP, respectively. High correlations among variates, a typical property of dynamic data, suggest that multivariate GP improves the prediction accuracy because MTG and TMGP can leverage the among-characteristics correlation. In the following, we focus on the comparison between MGP and TMGP.

In 2018, the accuracy of TMGP was lower than that of MGP in CV1 and CV2 (Figure 7) because of the low heritability of the LAI dynamics model parameters. However, the accuracy of TMGP was higher than that of MGP in 2019 for CV2. TMGP was better than MGP because of the higher heritability of growth parameters than GF in 2018. The extraction of genetic variance in growth patterns in 2019 was successful as LAI_{amp} in 2019-C and T_g in 2019-D, leading to improved prediction accuracy.

In CV3, the prediction accuracies of TGP and TMGP outperformed the other models (Figure 7). The higher prediction accuracy compared to MGP indicates that the former growth period's data could be effectively included in the model by specifying the growth curve's shape through the dynamics model. In most cases, the correlation coefficients between the predicted values of TMGP and the genotypic values exceeded 0.6, indicating that TMGP is robust to changes in the environment. The similar prediction accuracy of MGP and TMGP in 2017-D may be due to the lack of change in the GF in the second half of this environment's growth period. This approach to future prediction through dynamic models has potential applications for selection in early growth stages in crop breeding.

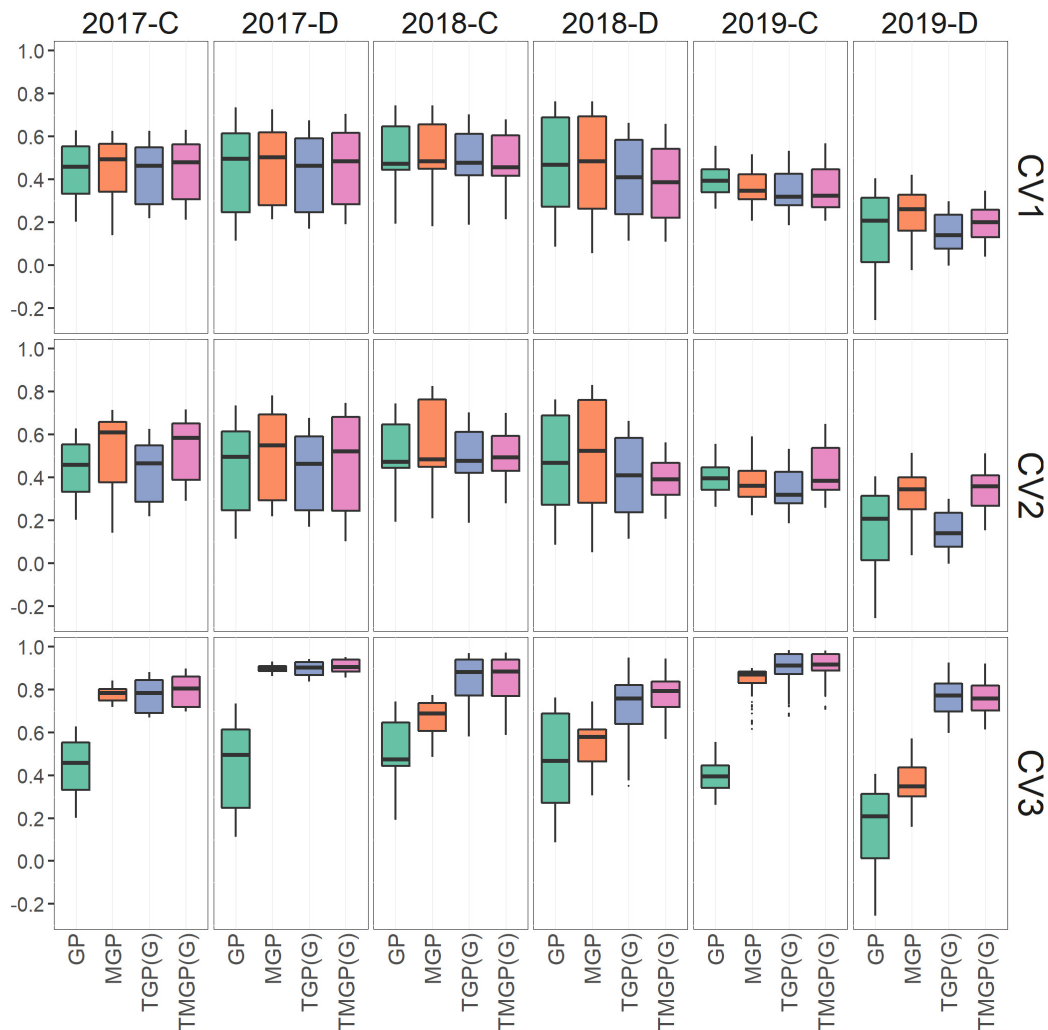


FIGURE 7 | Comparison of accuracy of the prediction models. The correlation coefficients between genotypic values and their predicted values are plotted. The results of CV1, CV2, and CV3 are shown in three panels.

In this study, the dynamics model and GP/MGP were used separately in TGP/TMGP, but they could be integrated into one hierarchical model. Several reports have shown the effectiveness of hierarchical models in the analysis of dynamic traits (Onogi et al., 2019), quantitative trait loci analysis (Ma et al., 2002), and genome-wide association studies (Das et al., 2011; Crispim et al., 2015). The joint analysis is expected to make parameter estimation more robust. In this study, although two steps are required to estimate the growth parameters, joint estimation may simplify the estimation process further.

Growth Analysis on Remote Sensing Data for Plant Breeding

Applying the dynamics model to crops monitored with UAVs allows us to capture the genetic variation in growth patterns. The combination of the dynamics model and genetic analysis was shown to be an efficient framework for analyzing our field experiments. It was able to predict future GF dynamics from

observations covering only the early growth stages. This could contribute to reducing the cost of the field trials. This study suggests that data monitoring the experiment with UAVs and analyzing them using dynamics models and mixed models will benefit crop breeding.

Although this study applied the growth model that considers both growth and senescence to soybean GF, characteristics of growth curves vary depending on species, trait, or situation of observation. For example, a logistic curve that consider only growth was used for modeling stem diameter of forest tree (Ma et al., 2002) and power function was used for modeling leaf age of rice (Wu et al., 2002). Crop models that consider physiological mechanisms such as photosynthesis may be applied to consider effects of diverse environmental factors on dynamic traits. Even in such cases, the proposed framework is flexible enough to be applied. In particular, the future prediction of growth curves (CV3) is a characteristic method of this framework and is expected to be applied to various traits.

A random regression model is also known as a regression method of dynamic data with a mixed model structure, which was used in the GP of dynamic traits (Sun et al., 2017; Campbell et al., 2018). The strength of random regression lies in its simple formation, but it cannot incorporate the growth curve structure like the dynamics model in exchange. Our prediction framework attempts to improve the accuracy of future predictions by considering the features of growth curves in the modeling.

In the near future, UAV-RS is expected to play an active role in plant breeding and provide growth trajectory data from multiple breeding programs. It will be possible for breeders and researchers to focus on new genotypes to select and develop new varieties suitable for the target environment. The integrated use of dynamic models and GP will be a useful method to effectively link growth process data with marker genotype data to improve genetic gain for genomic selection.

DATA AVAILABILITY STATEMENT

The datasets of canopy area, daily temperature, and growth parameters can be found in the GitHub at https://github.com/YT100100/ReferenceData_2021_Frontiers.

AUTHOR CONTRIBUTIONS

MT, MN, HisT, AK, and HI were involved in funding acquisition. AK prepared 198 accessions of soybean genetic resources. YT, GS, YO, YY, HirT, HidT, MT, HisT, AK, MN, TF, and HI conducted the field experiments. YT, GS, and HI conducted UAV remote sensing. HK-K prepared the genome-wide marker genotype data.

REFERENCES

- Araus, J., and Cairns, J. E. (2014). Field high-throughput phenotyping: the new crop breeding frontier. *Trends Plant Sci.* 19, 52–61. doi: 10.1016/j.tplants.2013.09.008
- Blanc, J., Dutartre, D., Tixier, M.-H., Weiss, M., Comar, A., Praud, S., et al. (2019). A high-throughput model-assisted method for phenotyping maize green leaf area index dynamics using unmanned aerial vehicle imagery. *Front. Plant Sci.* 10:685. doi: 10.3389/fpls.2019.00685
- Borra-Serrano, I., Swaef, T. D., Quataert, P., Aper, J., Saleem, A., Saeys, W., et al. (2020). Closing the phenotyping gap: high resolution UAV time series for soybean growth analysis provides objective data from field trials. *Remote Sens.* 12:1644. doi: 10.3390/rs12101644
- Cabrera-Bosquet, L., Crossa, J., Zitzewitz, J., Serret, M., and Araus, J. (2012). High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. *J. Integr. Plant Biol.* 54, 312–320. doi: 10.1111/j.1744-7909.2012.01116.x
- Calus, M. P., and Veerkamp, R. F. (2011). Accuracy of multi-trait genomic selection using different methods. *Genet. Select. Evol.* 43:26. doi: 10.1186/1297-9686-43-26
- Campbell, M., Walia, H., and Morota, G. (2018). Utilizing random regression models for genomic prediction of a longitudinal trait derived from high-throughput phenotyping. *Plant Direct* 2:e00080. doi: 10.1002/pld3.80
- Crispim, A. C., Kelly, M. J., Guimarães, S. E., Fonseca e Silva, F., Fortes, M. R., Wenceslau, R. R., et al. (2015). Multi-Trait GWAS and new candidate genes annotation for growth curve parameters in Brahman cattle. *PLoS One* 10:e0139906. doi: 10.1371/journal.pone.0139906
- YT, RL-L, and FB developed a dynamic model of the GF. YT analyzed the data and wrote the manuscript. HI supervised the study and edited the manuscript. All authors have reviewed and approved the final manuscript.
- Das, K., Li, J., Wang, Z., Tong, C., Fu, G., Li, Y., et al. (2011). A dynamic model for genome-wide association studies. *Hum. Genet.* 129, 629–639. doi: 10.1007/s00439-011-0960-6
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome J.* 4:250. doi: 10.3835/plantgenome2011.08.0024
- Endelman, J. B., and Jannink, J. -L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genet.* 2, 1405–1413. doi: 10.1534/g3.112.004259
- Furbank, R. T., and Tester, M. (2011). Phenomics – technologies to relieve the phenotyping bottleneck. *Trends Plant Sci.* 16, 635–644. doi: 10.1016/j.tplants.2011.09.005
- Jia, Y., and Jannink, J. -L. (2012). Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics* 192, 1513–1522. doi: 10.1534/genetics.112.144246
- Jin, X., Kumar, L., Li, Z., Feng, H., Xu, X., Yang, G., et al. (2018). A review of data assimilation of remote sensing and crop models. *Eur. J. Agron.* 92, 141–152. doi: 10.1016/j.eja.2017.11.002
- Kaga, A., Shimizu, T., Watanabe, S., Tsubokura, Y., Katayose, Y., Harada, K., et al. (2012). Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breed. Sci.* 61, 566–592. doi: 10.1270/jsbbs.61.566
- Kajiya-Kanegae, H., Nagasaki, H., Kaga, A., Hirano, K., Ogiso-Tanaka, E., Matsuoka, M., et al. (2021). Whole-genome sequence diversity and association analysis of 198 soybean accessions in mini-core collections. *DNA Res.* 28:dsaa032. doi: 10.1093/dnares/dsaa032
- Koetz, B., Baret, F., Poilvé, H., and Hill, J. (2005). Use of coupled canopy structure dynamic and radiative transfer models to estimate biophysical canopy

FUNDING

This study was supported by JST CREST (<https://www.jst.go.jp/kisoken/crest/en/index.html>) (grant no. JPMJCR16O2). The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We are grateful to the technical staff of the Arid Land Research Center, Tottori University, and Izumi Higashida for the management of the field experiments. We would like to thank Yuji Sawada for being involved in the field experiments, Kosuke Hamazaki for curating whole-genome marker data, and Sawako Maruyama for curating the phenotype data. This manuscript was published in agriRxiv (<https://agrirxiv.org>) prior to the submission in Frontiers in Plant Science (Toda et al., 2021b).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.828864/full#supplementary-material>

- characteristics. *Remote Sens. Environ.* 95, 115–124. doi: 10.1016/j.rse.2004.11.017
- Krishnappa, G., Savadi, S., Tyagi, B. S., Singh, S. K., Masthigowda, M. H., Kumar, S., et al. (2021). Integrated genomic selection for rapid improvement of crops. *Genomics* 113, 1070–1086. doi: 10.1016/j.ygeno.2021.02.007
- Liu, S., Baret, F., Abichou, M., Boudon, F., Thomas, S., Zhao, K., et al. (2017). Estimating wheat green area index from ground-based LiDAR measurement using a 3D canopy structure model. *Agric. For. Meteorol.* 247, 12–20. doi: 10.1016/j.agrformet.2017.07.007
- Ma, C.-X., Casella, G., and Wu, R. (2002). Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* 161, 1751–1762. doi: 10.1093/genetics/161.4.1751
- Madec, S., Baret, F., de Solan, B., Thomas, S., Dutartre, D., Jezequel, S., et al. (2017). High-throughput phenotyping of plant height: comparing unmanned aerial vehicles and ground LiDAR estimates. *Front. Plant Sci.* 8:2002. doi: 10.3389/fpls.2017.02002
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- Nelder, J. A. (1961). The fitting of a generalization of the logistic curve. *Biometrics* 17, 89–110. doi: 10.2307/2527498
- Onogi, A., Ogino, A., Sato, A., Kurogi, K., Yasumori, T., and Togashi, K. (2019). Development of a structural growth curve model that considers the causal effect of initial phenotypes. *Genet. Sel. Evol.* 51:19. doi: 10.1186/s12711-019-0461-y
- Piles, M., Gianola, D., Varona, L., and Blasco, A. (2003). Bayesian inference about parameters of a longitudinal trajectory when selection operates on a correlated trait. *J. Anim. Sci.* 81, 2714–2724. doi: 10.2527/2003.81112714x
- Soltani, A., and Sinclair, T. (2012). *Modeling Physiology of Crop Development, Growth and Yield*. Wallingford: CABI.
- Sun, J., Rutkoski, J. E., Poland, J. A., Crossa, J., Jannink, J.-L., and Sorrells, M. E. (2017). Multitrait, random regression, or simple repeatability model in high-throughput phenotyping data improve genomic prediction for wheat grain yield. *Plant Genome* 10, 1–12. doi: 10.3835/plantgenome2016.11.0111
- Toda, Y., Kaga, A., Kajiya-Kanegae, H., Hattori, T., Yamaoka, S., Okamoto, M., et al. (2021a). Genomic prediction modeling of soybean biomass using UAV-based remote sensing and longitudinal model parameters. *Plant Genome* 14:e20157. doi: 10.1002/tpg2.20157
- Toda, Y., Sasaki, G., Omori, Y., Yamasaki, Y., Takahashi, H., Takanashi, H., et al. (2021b). Genomic prediction of green fraction dynamics in soybean using UAV observations. *agRxiv* [preprint]. doi: 10.31220/agrxiv.2021.00097
- Verger, A., Vigneau, N., Chéron, C., Gilliot, J.-M., Comar, A., and Baret, F. (2014). Green area index from an unmanned aerial system over wheat and rapeseed crops. *Remote Sens. Environ.* 152, 654–664. doi: 10.1016/j.rse.2014.06.006
- White, J. W., Andrade-Sanchez, P., Gore, M. A., Bronson, K. F., Coffelt, T. A., Conley, M. M., et al. (2012). Field-based phenomics for plant genetics research. *Field Crops Res.* 133, 101–112. doi: 10.1016/j.fcr.2012.04.003
- Winsor, C. P. (1932). The gompertz curve as a growth curve. *Proc. Natl. Acad. Sci. U. S. A.* 18, 1–8. doi: 10.1073/pnas.18.1.1
- Wu, R., Ma, C.-X., Lin, M., and Casella, G. (2004). A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* 166, 1541–1551. doi: 10.1534/genetics.166.3.1541
- Wu, W., Zhou, Y., Li, W., Mao, D., and Chen, Q. (2002). Mapping of quantitative trait loci based on growth models. *Theor. Appl. Genet.* 105, 1043–1049. doi: 10.1007/s00122-002-1052-8
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., et al. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: current status and perspectives. *Front. Plant Sci.* 8:1111. doi: 10.3389/fpls.2017.01111

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 Toda, Sasaki, Ohmori, Yamasaki, Takahashi, Takanashi, Tsuda, Kajiya-Kanegae, Lopez-Lozano, Tsujimoto, Kaga, Nakazono, Fujiwara, Baret and Iwata. 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.



Full-Length Transcriptional Analysis of the Same Soybean Genotype With Compatible and Incompatible Reactions to *Heterodera glycines* Reveals Nematode Infection Activating Plant Defense Response

Minghui Huang^{1†}, Ye Jiang^{1,2†}, Ruifeng Qin^{1,2}, Dan Jiang^{1,2}, Doudou Chang^{1,2}, Zhongyan Tian², Chunjie Li^{1*‡} and Congli Wang^{1*‡}

OPEN ACCESS

Edited by:

Madan K. Bhattacharyya,
Iowa State University, United States

Reviewed by:

Peijian Cao,
Zhengzhou Tobacco Research
Institute of CNTC, China
Yan Li,
Sichuan Agricultural University, China

*Correspondence:

Chunjie Li
lichunjie@iga.ac.cn
Congli Wang
wangcongli@iga.ac.cn

[†]These authors have contributed
equally to this work and share first
authorship

[‡]These authors share last authorship

Specialty section:

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

Received: 31 January 2022

Accepted: 22 March 2022

Published: 18 May 2022

Citation:

Huang M, Jiang Y, Qin R, Jiang D,
Chang D, Tian Z, Li C and Wang C
(2022) Full-Length Transcriptional
Analysis of the Same Soybean
Genotype With Compatible
and Incompatible Reactions
to *Heterodera glycines* Reveals
Nematode Infection Activating Plant
Defense Response.
Front. Plant Sci. 13:866322.
doi: 10.3389/fpls.2022.866322

¹ Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Harbin, China, ² Heilongjiang Academy of Agricultural Sciences, Daqing, China

Full-length transcriptome sequencing with long reads is a powerful tool to analyze transcriptional and post-transcriptional events; however, it has not been applied on soybean (*Glycine max*). Here, a comparative full-length transcriptome analysis was performed on soybean genotype 09-138 infected with soybean cyst nematode (SCN, *Heterodera glycines*) race 4 (SCN4, incompatible reaction) and race 5 (SCN5, compatible reaction) using Oxford Nanopore Technology. Each of 9 full-length samples collected 8 days post inoculation with/without nematodes generated an average of 6.1 GB of clean data and a total of 65,038 transcript sequences. After redundant transcripts were removed, 1,117 novel genes and 41,096 novel transcripts were identified. By analyzing the sequence structure of the novel transcripts, a total of 28,759 complete open reading frame (ORF) sequences, 5,337 transcription factors, 288 long non-coding RNAs, and 40,090 novel transcripts with function annotation were predicted. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of differentially expressed genes (DEGs) revealed that growth hormone, auxin-activated signaling pathway and multidimensional cell growth, and phenylpropanoid biosynthesis pathway were enriched by infection with both nematode races. More DEGs associated with stress response elements, plant-hormone signaling transduction pathway, and plant-pathogen interaction pathway with more upregulation were found in the incompatible reaction with SCN4 infection, and more DEGs with more upregulation involved in cell wall modification and carbohydrate bioprocess were detected in the compatible reaction with SCN5 infection when compared with each other. Among them, overlapping DEGs with a quantitative difference was triggered. The combination of protein-protein interaction with DEGs for the first time indicated that nematode infection activated the interactions between transcription factor WRKY and VQ (valine-glutamine motif) to contribute to soybean defense. The knowledge of the SCN-soybean interaction mechanism as a model will present more understanding of other plant-nematode interactions.

Keywords: full-length transcriptome sequencing, *glycine max*, *Heterodera glycines*, incompatible and compatible response, VQ-WRKY interaction

INTRODUCTION

Soybean cyst nematode (SCN, *Heterodera glycines* Ichinohe) is one of the most economically important diseases in soybean (*Glycine max* L. Merrill) worldwide. The latest statistical analysis of estimated cumulative soybean economic losses due to diseases in 28 states within the United States from 1996 to 2016 indicated that SCN accounted for 23.2% of the total losses (US\$ 73,535 per hectare) which was top-ranked, far greater than the second disease charcoal rot (10%) (Bandara et al., 2020). In China, annual economic losses in soybean could reach up to US\$ 120 million (Li Y. et al., 2011). SCN is a soil-borne sedentary parasitic nematode in the major host soybean. In the only free-moving stage of SCN, a second-stage juvenile (J2) hatches from an egg (first-stage juvenile inside the egg) to search for host roots through signals released from plant roots, moves to root tips, and penetrates roots using a stylet. The J2 moves inside the root, establishes a feeding site, and reprograms host root cells by directing gland secretions into plants that can dissolve cell walls and fuse the protoplast of neighboring cells, and eventually forms a unique feeding structure called syncytium as a nutrient source for nematode development, consequently suppressing plant growth and affecting yield (Niblack et al., 2006; Mitchum and Baum, 2008).

Host-plant resistance combined with crop rotation is the most effective way to control SCN. More than 300 quantitative trait loci (QTLs) associated with SCN resistance were mapped to 20 chromosomes (chr) (soybase.org), but only two major resistance genes, *rhg1* on chromosome (chr) 18 and *Rhg4* on chr 8, were cloned and characterized (Cook et al., 2012, 2014; Liu et al., 2012, 2017). The presence of multiple minor QTLs contributing to resistance in either resistant or susceptible soybean plants makes resistance breeding more difficult than expected (Huang et al., 2021). Furthermore, shifts in virulence have caused a decrease or loss of resistance because of the long-term planting of a single source of resistance varieties, e.g., 90% of resistance sources derived from the PI88788 background in the United States and mainly Peking resistance sources in China (Mitchum et al., 2007; Niblack et al., 2008; Acharya et al., 2016; Hua et al., 2018; Huang et al., 2022). Furthermore, Peking and PI88788 display resistance only to some SCN races or HG Types. The lack of broad resistance sources and the presence of multiple SCN races or HG types in the field result in SCN spreading widely and quickly. Thus, understanding the molecular mechanisms of SCN infection and plant resistance will gain more insights to develop new control strategies, including engineering important candidate genes to increase resistance.

Plants have evolved to develop two layers of pathogen defense immune systems: the first layer is pathogen- or microbe-associated molecular pattern (PAMP/MAMP) triggered immunity (PTI); the second layer is effector-triggered immunity (ETI), which fits species-specific disease resistance (Dangl and Jones, 2001; Eitas and Dangl, 2010; Monaghan and Zipfel, 2012). The plant cell wall surface contains pattern recognition receptors (PRRs) that can detect pathogen or microbe structures to activate PTI, and phytohormones as defense-related signaling

molecules, such as salicylic acid (SA), jasmonate acid (JA), ethylene (ET), induce plant to produce pathogenesis-related (PR) proteins, e.g., β -1,3-glucanases, peroxidase, oxidase-like, thionin, and proteinase inhibitor, etc. (Sels et al., 2008). ETI is initiated by intracellular nucleotide-binding protein domain leucine-rich repeat proteins (NLRs) to generate a hypersensitive response (HR) with local cell death (Eitas and Dangl, 2010). PTI and ETI are activated by two distinct classes of receptors in early signaling, but recent evidence suggests that PTI and ETI can crosstalk in downstream response, although how they contribute to immunity with quantitative and/or qualitative outputs is still undefined (Naveed et al., 2020; Yuan et al., 2021). Transcriptome analysis of different pathosystems denotes that both compatible and incompatible interactions are able to trigger an overlapping change of gene expression but with quantitative differences (Mine et al., 2018; Yuan et al., 2021).

Studies have been performed on compatible and incompatible interactions between SCN and soybean roots based on microarray and RNA-seq transcriptome analyses (Ithal et al., 2007; Klink et al., 2007; Puthoff et al., 2007; Klink and Matthews, 2009; Kandoth et al., 2011; Li X. et al., 2011; Li et al., 2012, 2018; Mazarei et al., 2011; Wan et al., 2015; Zhang et al., 2017; Kang et al., 2018; Neupane et al., 2019; Song et al., 2019; Jiang et al., 2020; Miraeiz et al., 2020). All sequence annotations revealed that SCN infection can induce or suppress gene expression either in susceptible or resistant cultivars, and a series of defense genes (*PPRs* and *NLRs*), *MAPK* (mitogen-activated protein kinase) signaling cascade, *WRKY* and *MYB* transcription factors (TFs), heat shock protein (*HSP*) genes, *PR* genes, and phenylpropanoid metabolism genes have been identified but with variance depending on SCN race/HG type and plant type.

RNA-sequencing (RNA-seq) based on high-throughput next-generation sequencing (NGS) (e.g., Illumina) has been used widely to measure differential gene expression because it is a cost-effective and advanced technology (Finotello and Di Camillo, 2015). However, RNA-seq requires fragmentation of RNA or cDNA to generate short reads when preparing samples, which diminishes the information from original full-length transcripts; thus, it is harder to obtain post/co-transcriptional processing events that are responsible for producing a mature RNA molecule that can leave the nucleus to function in the cell by chemical structure alteration of the RNA primary transcript (Kiss, 2001). Advanced full-length transcriptome sequencing with longer reads circumvents these challenges. Currently, PacBio and Oxford Nanopore (ONT) are the most popular 3rd-generation full-length sequencing technologies that can provide more complex transcription and reveal the real structure of sequences during transcription, such as alternative splicing (AS), alternative polyadenylation (APA), and long non-coding RNA (lncRNA) and gene fusion, which can increase the complexity of the transcriptome and proteome. Compared with PacBio sequencing, ONT uses ion current blockades to directly sequence more long-native DNA or full-length RNA molecules (Cui et al., 2020; Xie et al., 2021). AS, one of the important steps in post-transcriptional modification, can recognize and eliminate intronic regions of a precursor messenger RNA (pre-mRNA) to generate multiple mRNAs to regulate gene expression that consequently promotes

proteome diversity. AS plays key roles not only in plant growth and development but also in response to biotic or abiotic stimuli or adaptation (Matsukura et al., 2010; Severing et al., 2012; Syed et al., 2012; Mandadi and Scholthof, 2015; Wang et al., 2018; Bedre et al., 2019; Martín et al., 2021). APA can produce multiple mRNA polyadenylation isoforms through pre-mRNA endonucleolytic cleavage and poly(A) tail addition at the 3' cleavage site end of a nascent transcript to change the length of untranslated regions (UTRs) or coding regions that may influence mRNA stability, translation efficiency, subcellular localization, or gene function gain or loss. Consequently, all these changes will result in various plant physiological and biochemical processes (Yeh and Yong, 2016; Sadek et al., 2019; Zhang et al., 2020; Tu et al., 2021); for example, APA participates in cell wall modification, root hair development, DNA repair, and gene regulation in response to abiotic and biotic stresses (Cao et al., 2019; Ye et al., 2019; Yan et al., 2021). LncRNAs longer than 200 nucleotides are epigenetic regulators that regulate gene expression by interacting with mRNAs, DNAs, proteins, and miRNAs to participate in biological processes such as plant growth and development and biotic and abiotic stress responses (Budak et al., 2020; Yu et al., 2020; Tu et al., 2021; Urquiaga et al., 2021).

The application of the full-length transcriptome sequencing technique in plants is limited compared with second-generation sequencing because of its higher cost. Currently, full-length transcriptome sequencing for inferring and improving gene models and identifying novel genes has been reported on rice, wheat, maize, cotton, pecan, poplar, and others but not on soybean (Clavijo et al., 2017; Wang et al., 2018; Zhang et al., 2019; Zhao et al., 2019; Li C. et al., 2020; Yang et al., 2021).

Our previous study has demonstrated that the soybean breeding line 09-138, developed in northeast China, carries the *rhg1-a* and *Rhg4-b* loci and has resistance to SCN race 4 (SCN4, HG type 1.2.3.5.6.7) but has susceptibility to SCN race 5 (SCN5, HG type 2.5.7), which is different from the Peking (*rhg1-a* + *Rhg4-a*) and PI88788 (*rhg1-b* + *Rhg4-b*) resistance backgrounds (Hua et al., 2018; Huang et al., 2022). We postulated that differences in response to the two SCN races should be related to distinct transcriptional responses in the early stages of nematode infection as described above. To test this, line 09-138 was inoculated with the two SCN races, and nematode development was observed inside the roots. Comparative full-length transcriptome analysis of soybean breeding lines 09-138 infected with SCN4 (resistance response) and SCN5 (susceptibility response) was performed using the ONT technology. We investigated AS and APA events and identified lncRNA and transcription factors in detected novel transcripts. Then, differentially expressed genes (DEGs) and transcripts (DETs) were analyzed, and DEGs associated with stress response elements were explored. Enriched DEG-GO (Gene Ontology) terms and DEG-KEGG (Kyoto Encyclopedia of Genes and Genomes) pathways were compared between resistant and susceptible responses. Protein–protein interactions were predicted, and a defense mode was established. Finally, the expression of DEGs was validated by quantitative RT-PCR. Comparisons between compatible and incompatible responses

will provide insight into the resistance mechanism and identify candidate defense or resistance genes for further study.

MATERIALS AND METHODS

Plant Materials and Nematode Culture

The soybean breeding line 09-138 was developed by the Heilongjiang Academy of Agricultural Sciences (Hua et al., 2018). SCN race 4 (SCN4, HG type 1.2.3.5.6.7) and SCN race 5 (SCN5, HG type 2.5.7) were originally collected from the field and cultured from single cysts for more than 5 generations on the susceptible soybean variety Dongsheng1 in a greenhouse with 16 h of light and 8 h of darkness at 23–28°C, and then identified by race test and HG type indicator lines (Hua et al., 2018). Every other year, nematode HG types/races were reconfirmed without virulence change with time.

A nematode inoculum was prepared according to the method described by Huang et al. (2022). Plant root tissue and soil were collected 35–40 days after inoculation and put into a 2 L-beaker. The mixture was stirred vigorously with a glass rod for approximately 1 min and then precipitated for 10 s. The fluid of the supernatant was gently poured into 75/25 µm nested sieves. The mixture of cysts and root debris on the top of the sieves was rubbed with a rubber stopper to release eggs. Eggs on 25-µm sieve were rinsed with a high-pressure water faucet for 1 min and then with sterile water before collection. The collected eggs were then transferred onto six to eight layers of tissue paper supported by a metal screen on a hatching dish containing 3 mM ZnSO₄ in sterile water for hatch at 28°C. J2s were then collected for inoculation after 3–4 days.

Nematode Inoculation, Root Staining, and Root Preparation for RNA Extraction

Seeds of 09-138 were sterilized by soaking in 0.5% sodium hypochlorite for 20 min and rinsed with sterile water three times. Two seeds were sown in a black plastic pot (8 cm diameter × 12 cm depth) filled with autoclaved soil and sand at a ratio of 1:1. After 4 days, two seedlings were thinned to one in each pot. An eight-day seedling was inoculated with a 1-ml suspension containing 2,000 J2 of SCN4 or SCN5. Seedlings were inoculated with 1 ml water as control. The plants were maintained in a growth chamber at a 16-/8-h day/night regime, 28°C day/22°C night, and 50% relative humidity.

Collected roots were stained 3, 6, 8, 10, and 12 days after inoculation with acid fuchsin (Byrd et al., 1983). Nematode development inside the roots was observed, and roots and nematodes were photographed under an Olympus SZX16 dissecting microscope using the Cellsens Standard image software (Olympus Corporation, Japan).

To collect roots for RNA extraction, plant roots 8 days after inoculation were washed and rinsed thoroughly with water. Three roots from each treatment were wrapped together with aluminum foil as one replication (one sample). Three replications were made for each treatment. A total of 9 samples with SCN4- and SCN5-infection and control were collected. Immediately, each prepared sample was put in liquid nitrogen to freeze it and was kept at -80°C for RNA extraction and sequencing.

RNA Extraction, cDNA Library Construction, and Nanopore Sequencing

Total RNA was extracted using RNAeasy Plant Mini Kit (Qiagen, United States), and RNase-free DNase (Qiagen) was used to remove DNA contamination in the total RNA. The concentration, purity, and integrity of the extracted RNA were measured with 1% agar gel (Thermo Fisher Scientific, United States) and Agilent 2100 Bioanalyzer (Agilent Technologies, United States). cDNA library construction started with 1 µg total RNA using a cDNA-PCR sequencing kit (SQK-PCS109) provided by Oxford Nanopore Technologies (ONT, Inc., United Kingdom) following the manufacturer's instructions. Final cDNA libraries were added to FLO-MIN109 flow-cells and run on the PromethION platform at Biomarker Technology Company (Beijing, China) for sequencing. Experimental processes including sample quality testing, library building, library-quality testing, and library sequencing were performed in accordance with standard procedures provided by ONT.

Raw Data Processing to Obtain Full-Length Transcriptome

Raw reads were subject to filtering with an average read quality score ≤ 7 , and read length ≤ 500 bp, and ribosomal RNA mapped to the rRNA database was also discarded. Primer sequences on both ends of clean reads were searched to determine full-length, non-chimeric (FLNC) sequences. Detected FLNC transcripts were then mapped to soybean reference genome Williams 82.a2.v1¹ with minimap2 (Li, 2018) to obtain FLNC clusters, and pinfo² was applied to polish each cluster to attain consensus isoforms. All mapped reads were further collapsed using the cDNA_Cupcake package with a mini-identity of 90% and mini-coverage of 85%. When the redundant transcripts were collapsed, 5' difference was not considered. The achieved transcripts were compared with known transcripts of reference genome Williams 82.a2.v1 utilizing gffcompare, and novel transcripts were identified in order to make supplementary genome annotation. Gene boundaries were modified, and transcripts with expressed levels ≤ 1 were filtered.

Structure Analysis: Identification of Alternative Polyadenylation, Fusion Transcript, Alternative Splicing Events, and Microsatellite Markers

Alternative polyadenylation was identified through further analysis of FLNC by Transcriptome Analysis Pipeline from Isoform Sequencing (TAPIS) (Foissac and Sammeth, 2007). Multiple Expectation Maximization for Motif Elicitation (MEME) (Bailey et al., 2006) was used to analyze the 50-bp sequence upstream of the poly A site to detect FL motifs. The consensus sequence before remove-redundant analysis was used for fusion transcript analysis. Fusion transcript was defined under the following conditions: aligned to 2 or more sites; each site covers at least

5% of a transcript with a 1-bp minimum alignment length; the total length covers more than 95% of the total length of transcripts with at least 10k bp distance between the two sites.

Alternative splicing indicates the process of pre-mRNA treatment. Gene transcription generates pre-mRNAs with many splicing methods. Five types of alternative splicing events (3' splice site, 5' splice site, exon skipping, intron retention, and mutually exclusive exon) of transcripts were examined by employing the Astalavista software based on the alignment results of individual samples to the reference genome (Foissac and Sammeth, 2007). The MicroSatellite (MISA, an identification tool) software was used for SSR analysis, and transcripts below 500 bp were discarded.

Coding Sequence Prediction, Long Non-coding RNA Identification, and Transcription Factor Detection From Novel Transcripts

Coding sequences (CDSs) were predicted with TransDecoder (v3.0.0; Haas et al., 2013) based on the ORF. LncRNA does not code for protein. Therefore, LncRNA in novel transcripts was predicted whether it had a coding potential by protein domain analysis including all four methods, Coding Potential Calculator (CPC) (Kong et al., 2007), Coding-Non-Coding Index (CNCI) (Sun et al., 2013), Coding Potential Assessment Tool (CPAT) (Wang et al., 2013) and Protein family (Pfam) (Finn et al., 2014). LncRNA target genes were predicted using two methods: first, depending on the location relationship between the differentially expressed lncRNA and adjacent mRNA (within 100k bp distance) expressed differentially; second, according to complementary base pairing between lncRNA and mRNA using the lncTAR tool (Li et al., 2015). TFs were detected with iTAK (Zheng et al., 2016).

Quantification of Transcript/Gene Expression Levels and Differential Expression Analysis

Mapped full-length reads with > 5 match quality were chosen for quantification. Transcript or gene expression levels were measured in counts per million (CPM) (Zhou et al., 2014) and calculated by the following:

$$\text{CPM} = (\text{read number matched the transcript}) / (\text{total read number matched referenced transcriptome}) \times 10^6$$

The differential expression among the treatments was analyzed with DESeq2 (Anders and Huber, 2010) depending on a negative binary distribution model, consequently gaining DEGs or DETs. False discovery rate (FDR) was adjusted and controlled with the method of Benjamini and Hochberg (1995), and DEGs or DETs with \log_2 fold change (FC) ≥ 2 and FDR < 0.01 were chosen. A heat map for DEGs in each group was developed using the pheatmap package in R (Version 1.0.12³).

¹<https://phytozome.jgi.doe.gov/>

²<https://github.com/nanoporetech/pinfo>

³<https://CRAN.R-project.org/package=pheatmap>

Functional Annotation and Enrichment Analysis of Differentially Expressed Genes/Transcripts

Functional annotation of the genes/transcripts was conducted by blasting with databases including NR (NCBI non-redundant protein sequences) (Deng et al., 2006), Swissprot (Apweiler et al., 2004), GO (Ashburner et al., 2000), Clusters of Orthologous Groups (COG) (Tatusov et al., 2000), euKaryotic Ortholog Groups (KOG) (Koonin et al., 2004), Pfam (Kanehisa et al., 2004), and KEGG (Mckenna et al., 2010).

Gene ontology enrichment analysis of DEGs or DETs was conducted using the Goseq R package-based Wallenius non-central hypergeometric distribution (Young et al., 2010). KEGG pathway enrichment analysis of DEGs or DETs was subject to the KEGG Orthology Based Annotation System (KOBAS) software (Mao et al., 2005). Protein–protein interactions (PPIs) for all detected DEGs were predicted using the STRING database⁴ and were visualized in Cytoscape (Shannon et al., 2003).

Validation of Differentially Expressed Genes by Real-Time Quantitative Reverse Transcription-PCR

Samples leftover from full-length transcriptome sequencing was subject to qRT-PCR for verification of DEGs. Primers were designed by using the Primer Premier 5 software (Lalitha, 2000) and synthesized by Comate Bioscience Company Limited (Changchun, China). A total of 1 µg of treated RNA was used to synthesize the first-strand cDNA using FastKing gDNA Dispelling RT SuperMix (TIANGEN, China). PCR reactions were conducted in LightCycler[®] 480 System (Roche Life Science, United States) with ChamQ Universal SYBR qPCR Master Mix (Vazyme Biotech Co., Ltd., Nanjing, China) following the manufacturer's protocol. PCR was carried out in a 20-µl volume containing 100 ng cDNA (2 µl). PCR was performed as follows: initial denaturing for 10 min at 95°C, followed by 40 two-step cycles of 95°C for 10 s and then at 60°C for 1 min. The relative expression of tested genes was calculated with the $-\Delta\Delta C_t$ method using ACTIN as a control. Three independent biological replicates and three technical repetitions were performed for all the experiments. The primers used are listed in **Supplementary Table 1**. The correlation between transcriptome sequencing and qRT-PCR was completed in Excel 2016, and independent-samples *t*-test was performed using SPSS 17.0.

RESULTS

Differential Development of *Heterodera glycines* Race 4 and Race 5 Inside Roots of Lines 09-138 in Early Stages

Various development stages of the two races, SCN4 (**Figure 1A**) and SCN5 (**Figure 1B**), were observed in roots. Soybean roots were penetrated by both nematode races, and there was no

obvious difference in nematode size on day 3 days (**Figure 1**). On day 6, most of SCN5 developed to the J3 stage but SCN4 remained in the J2 stage. Nematodes had developed from J3 to late J4 stages in roots infected with SCN5 at 8, 10, and 12 days (**Figure 1B**) when compared with J2, J3, or a few early J4 stages in roots infected with SCN4 (**Figure 1A**), confirming 09-138 is resistant to SCN4 (cyst number/plant, $13 \pm SE 2.7$; female index, FI = 10) and susceptible to SCN5 (cyst number/plant, $119 \pm SE 8.16$; FI = 40) (Huang et al., 2022). More brown spots (hypersensitive response) around some nematode feeding sites were observed in resistant roots with SCN4 than in those with SCN5 (**Figure 1A**).

Nanopore Full-Length Transcriptome Sequence Statistics and Redundant Remove of Transcript

Since an obvious difference in nematode development on day 8 was observed in 09-138, root samples from day 8 were used for full-length sequencing. An average of 6.1 Gbp of clean data for nine cDNA libraries was obtained with a range from 5.73 to 6.82 Gbp (**Supplementary Table 2**). The average N50 length was 1,287 bp with a mean length of 1,142 bp, an average max length of 11,276 bp (9,575–13,426 bp), and a mean quality value of 11 (Q11) (**Supplementary Table 2**). After rRNA was filtered out, an average of 5,296,377 clean reads (4,681,541–5,888,283) and an average of 4,258,467 FLNC read numbers (3,767,343–4,641,061) were generated with an average of 80.4% FLNC ratio (**Table 1**).

Finally, 65,038 redundant-removed transcript sequences containing 92,079,818 bp with an N50 length of 1,665 bp, mean length of 1,415 bp, and max-length of 7,285 bp were obtained through a merging of the consistent sequences (**Supplementary Figure 1**). After alignment with the reference genome and filtering with known annotations of the reference genome, 1,117 novel genes and 41,096 novel transcripts were identified. The consistent sequence of each sample was used for AS analysis.

Structure Analysis of Alternative Polyadenylation Events, Transcript Fusion, Alternative Splicing Events, and SSR Prediction Revealed Soybean Structure Variation in Response to *Heterodera glycines* Race 4 and Race 5 Infections

Full-length sequencing can accurately identify the structure of transcripts. APA analysis based on FLNC displayed a total of 214,760 transcripts with various numbers of poly A sites. Among the three treatments, the greatest portion of 23.8% transcripts contained > 5 poly A sites, followed by 1 poly A site (21.5%), and the least with 5 poly A sites (9.6%) (**Figure 2A**). The number of transcripts in each poly A site distribution showed no significant difference ($P > 0.05$) among the three treatments (**Figure 2B**). In a more dissecting way, we found that SCN4 infection made soybean generate a significantly greater ($P < 0.05$) mean number ($840 \pm SE 26$) of >10 poly A sites than SCN5 infection ($734 \pm SE 12$) and the control ($698 \pm SE 26$), indicating APA with more poly A sites might be involved in the

⁴<http://string-db.org>

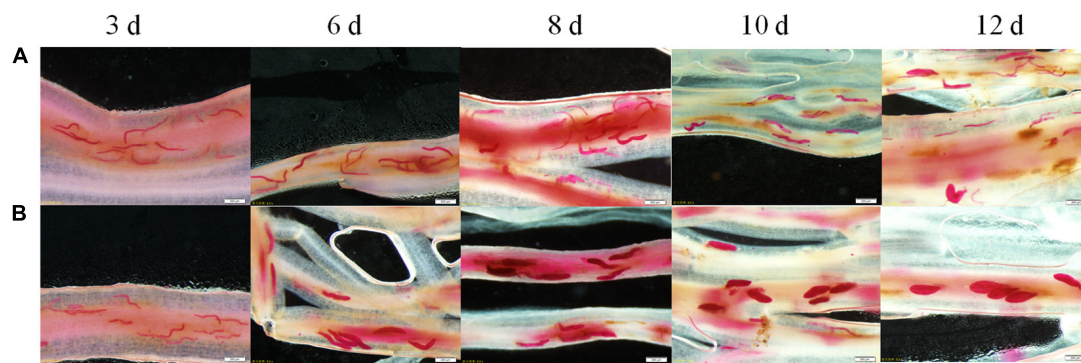


FIGURE 1 | Development of soybean cyst nematode **(A)** race 4 (HG type 1.2.3.5.6.7) and **(B)** race 5 (HG type 2.5.7) inside the roots of soybean genotype 09-138 at 3, 6, 8, 10, and 12 days as indicated at the top of the image after inoculation with 2,000 J2. The roots were stained with acid fuchsin. Scale bar = 200 μ m.

TABLE 1 | Number of clean reads and full-length reads, and full-length percentage.

Treatment ^a	Number of clean reads (except rRNA) ^b	Number of full-length reads ^c	Full-length percentage(%) ^d
CK-1	5,592,903	4,437,097	79.33
CK-2	5,683,524	4,641,061	81.66
CK-3	5,399,629	4,311,944	79.86
SCN4-1	5,459,017	4,409,582	80.78
SCN4-2	5,888,283	4,631,268	78.65
SCN4-3	5,184,003	4,146,135	79.98
SCN5-1	4,727,932	3,785,053	80.06
SCN5-2	5,050,558	4,196,724	83.09
SCN5-3	4,681,541	3,767,343	80.47
Mean	5,296,377	4,258,467	80.43

^aCK-1, CK-2, and CK-3 represent control with water for three biological replications on soybean breeding line 09-138; Similarly, SCN4-1, SCN4-2, and SCN4-3 represent root samples infected with soybean cyst nematode race 4 (HG type 1.2.3.5.6.7) for three replications; and SCN5-1, SCN5-2 and SCN5-3 represent SCN race 5 for three replications. ^bNumber of clean reads (except rRNA): number of clean read sequences after filtering rRNA. ^cNumber of full-length reads: number of full-length sequences. ^dFull-length percentage (%): percentage of number of full-length reads compared with clean reads.

incompatible reaction. Enrichment in T and A were detected in the upstream and downstream of the 50-bp position, respectively (Figure 2C). Three GC-rich APA motifs (CAGGGG, GGCTGC, and GGCCGC) were identified in the 50-bp sequence upstream of the poly A sites (Figure 2D). Fusion transcript screening before the remove-redundant analysis revealed 2–12 fusion transcripts for each sample. Similarly, SCN4 infection generated a total of 21 fusion transcripts, while only 14 fusion transcripts were found for SCN5 and the control (Supplementary Table 3), suggesting that fusion transcripts might be participating in defense response.

The Astaravista analysis of novel transcripts demonstrated that all the three treatments had no significant difference in the average number of AS events, $1,925 \pm \text{SE } 54.6$, $2,265 \pm \text{SE } 145.8$; and $2,287 \pm \text{SE } 130.1$ for the control, SCN4, and SCN5 treatments, respectively (Figure 2E), but that the AS events varied among the treatments. Based on the average percentage of AS events among the 9 libraries, there were 40.11% intron retentions, 26.28% alternative 3' splice sites, 18.81% exon skipping, 14.34% alternative 5' splice sites, and 0.46% mutually exclusive exons (Figure 2F). The numbers of AS events at alternative 3' splice sites in both nematode-infected treatments were significantly lower ($P < 0.05$) than those in the control, while the numbers

at 5' splice sites were significantly greater ($P < 0.05$) than those in the control.

A total of 22,486 SSRs were identified from 61,900 novel transcripts containing a total of 90,765,603 bp (Supplementary Table 4). Three types of SSR with mono-, di-, and tri-nucleotide accounted for 98% of total SSRs. There were 1,718 SSRs present in compound formation (hybrid microsatellite, a distance of two SSRs < 100 bp) (Supplementary Figure 2 and Supplementary Table 4).

Thus, all the tested structure variations among the control, SCN4, and SCN5 treatments demonstrated that post-transcriptional modification might be involved in nematode resistance or susceptibility in soybean.

Coding Sequences Prediction of Novel Transcripts

A total number of 37,469 ORFs were obtained, including 28,759 complete ORFs. Approximately, 38 and 43% of all predicted CDS coding protein lengths were within 0–100 and 100–200 aa, respectively (Figure 3A). The corresponding complete CDS protein lengths are shown in Figure 3A.

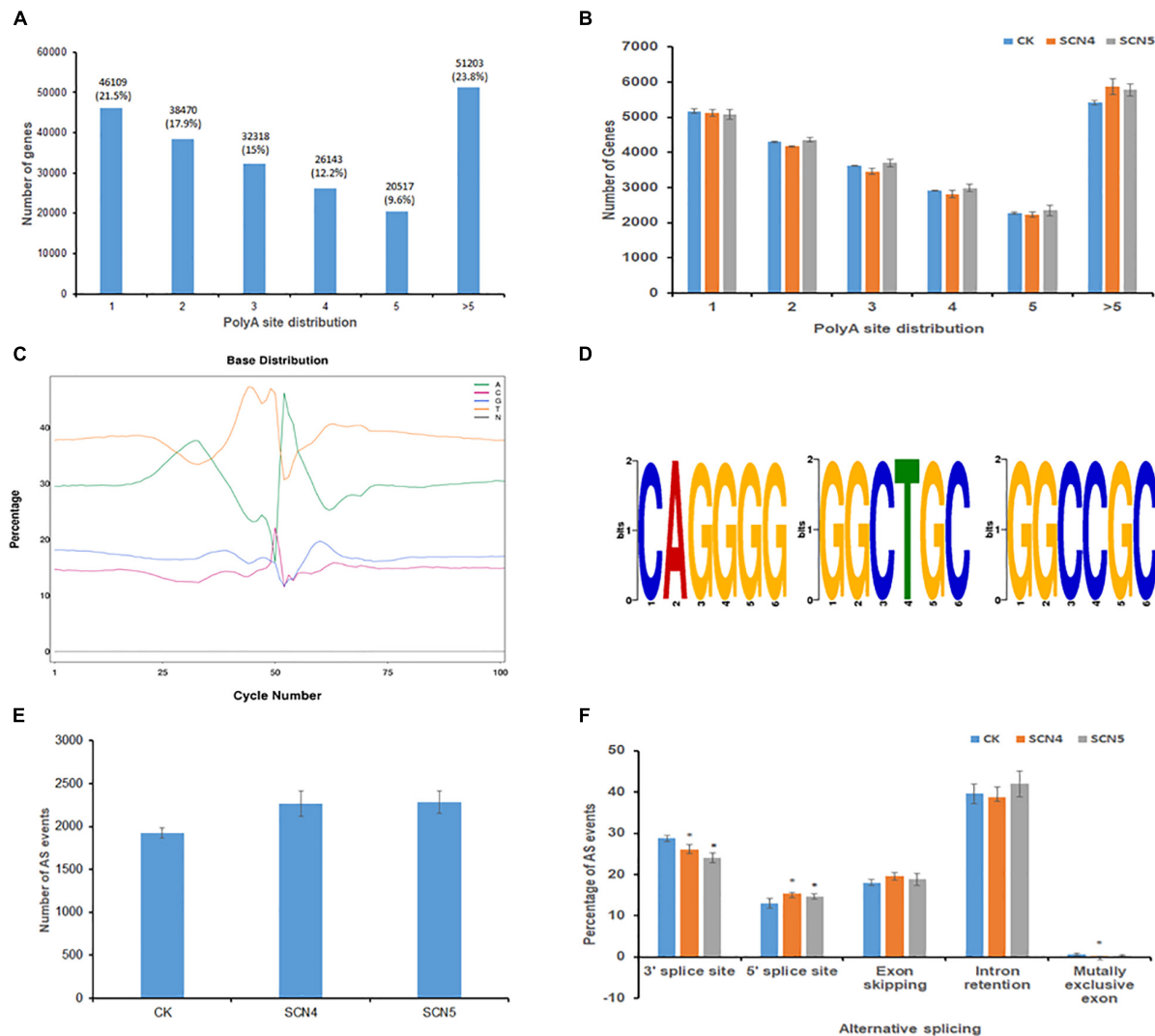


FIGURE 2 | Alternative polyadenylation (APA) and alternative splicing (AS) event analysis on soybean genotype 09-138 infected with soybean cyst nematode race 4 (SCN4) and race 5 (SCN5), and water as the control (CK). **(A)** Distribution of total poly A sites of genes. **(B)** Comparison of poly A site distribution among CK, SCN4, and SCN5 treatments. **(C)** Base distribution (in%) at 50 bp upstream and downstream of the poly A site of all transcripts. **(D)** Identified APA motifs. **(E)** Total AS number for CK, SCN4, and SCN5. **(F)** Comparison of AS events among CK, SCN4, and SCN5 treatments.

Long Non-coding RNA Associated With Defense Responsive Transcription Factors

Long Non-coding RNAs associated with plant growth, development, and stress responses have garnered widespread attention (Szcześniak et al., 2015; Urquiza et al., 2021). In total, 288 lncRNAs were predicted by CPC, CNC, CPAT, and Pfam protein domain analyses; 24 and 90 of them were unique to the CPAT and CPC methods, respectively (Figure 3B); 47.6 (137) and 42.4% (122) of them were classified as sense_lncRNA and lincRNA (long intergenic non-coding RNA), respectively (Figure 3C). Antisense-lncRNA and intronic-lncRNA accounted for 7.3 (21) and 2.8% (8), respectively (Figure 3C). The number of cis-targeted genes (275) regulated by these lncRNAs was

nearly four times greater than those (85) of trans-targeted genes. A lncRNA-Pfam protein domain analysis exhibited 164 transcripts matched to the Pfam database, including defense-responsive transcription factors WRKY, TIR, EF-hand, AUX/IAA, zf-RVT (zinc finger in the reverse transcript), zf-CCHC, BoA (bacterial stress-induced morphogen protein) and others, suggesting that lncRNAs may be involved in nematode stress response in soybean. Interestingly, 69 of 164 (42%) Pfam domains were extensin-like protein repeats from two novel transcripts, ONT.12398 (chr 9) and ONT. 12400 (chr 9), and 3 domains were extensin-like regions in ONT.14325 (chr 10). Extensins are a family of hydroxyproline-rich glycoproteins (HRGPs) in plant cell walls that can mediate resistance to viruses, bacteria, fungi, and nematodes (Deepak et al., 2010; Hirao et al., 2012). Especially, extensins identified in SCN syncytium

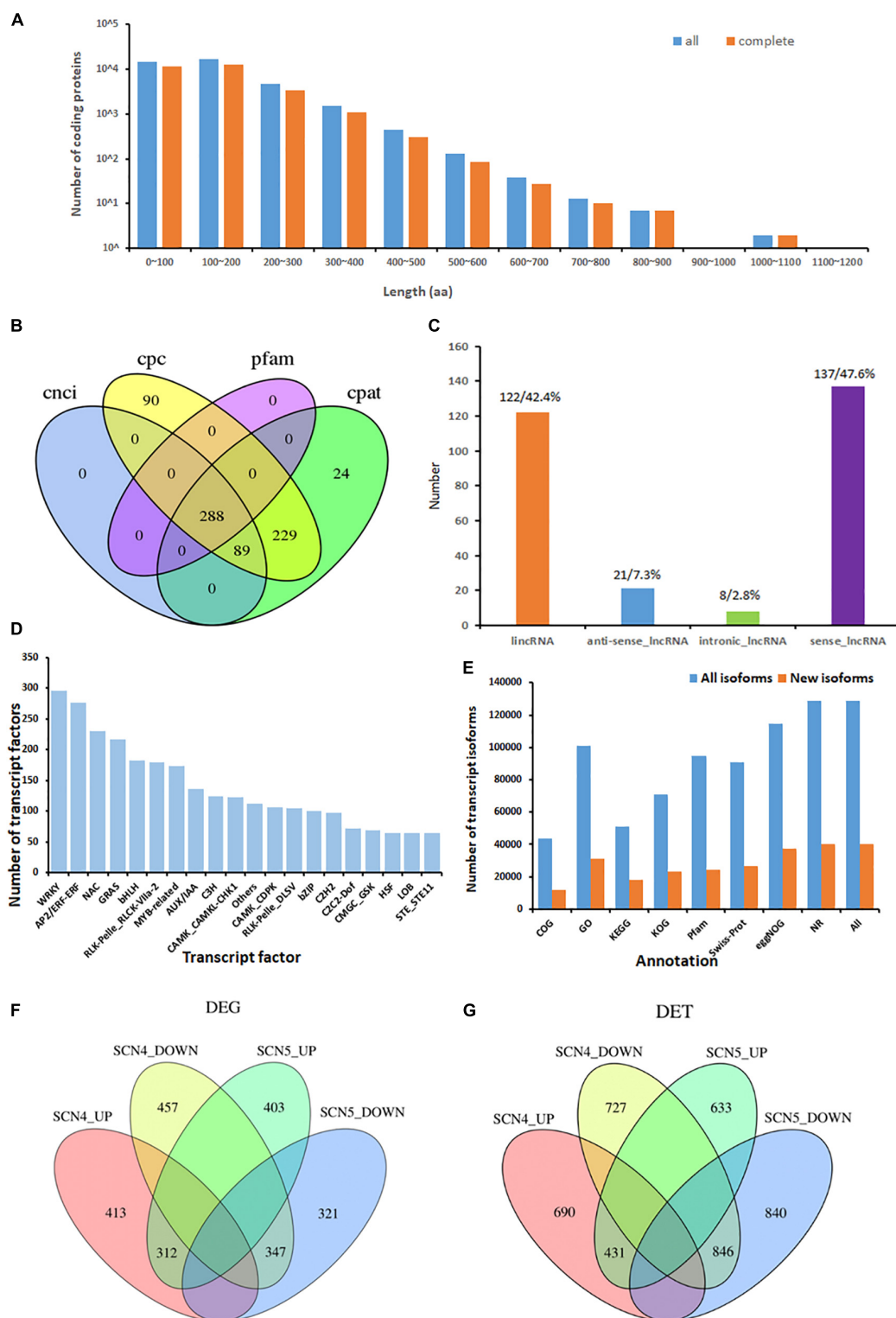


FIGURE 3 | Isoform coding sequence length, lncRNA, and transcription factors from the novel transcripts identified from 9 soybean samples, and differentially expressed genes (DEGs) or transcripts (DETs) in soybean response to soybean cyst nematode race 4 (SCN4) and race 5 (SCN5) when compared with the control (CK). **(A)** Length distribution of all/complete coding protein sequences predicted (unit, aa, amino acid). **(B)** Number of lncRNAs identified with the CPC, CNCI, CPAT, and Pfam methods. **(C)** Percentage of the classification of lncRNA positions annotated with the reference genomes. **(D)** Distribution of top 20 enriched transcription factor types. **(E)** Number of all and novel transcript isoforms annotated by the COG, GO, KEGG, KOG, Pfam, Swiss-Prot, eggNOG, and NR databases. **(F,G)** Number of upregulated and downregulated DEGs and DETs between the CK-SCN4 and CK-SCN5 treatments.

formation (Ithal et al., 2007; Matsye et al., 2011) and nematode-related lncRNA co-expression with HRGP potentially mediating SCN infection (Khoei et al., 2021) consistently confirmed that altered cell wall composition with extensins may play roles in disease defense.

Transcription Factor Prediction and Function Analysis of the Novel Transcripts

There were 5,337 TFs predicted, including 176 families of TFs (transcription factor, transcription regulator, and protein kinase) with a number ranging from 1 to 296 (**Supplementary Table 5**). The three treatments, the control, SCN4 and SCN5, contained 1,804, 1,787, and 1,746 TFs, respectively. The most abundant TFs were WRKY (296), AP2/ERF-ERF (APETALA2/Ethylene Responsive Factors, 276), NAC (NAM, ATAF, and CUC, 230), and GRAS (GAI, RGA, and SCA, 217) (**Figure 3D**).

Function annotation of transcripts obtained from the alternative splicing analysis exhibited 128,648 known and 40,090 novel transcript isoforms (**Supplementary Table 6**) and 56,865 known and 859 novel genes (**Supplementary Table 7**). The number of annotated transcripts in all and new isoforms is displayed in **Figure 3E**. Among them, NR annotation of species distribution indicated that 86.8, 8.9, 0.95, and 0.4% of the transcripts were aligned to *G. max*, *G. soja*, *Phaseolus vulgaris*, and *Medicago truncatula*, respectively. The GO-enrichment analysis showed that 13,483 out of 100,994 (13.5%) *G. max* isoforms and 4,350 out of 31,102 (14.0%) novel isoforms were associated with response to stimulus (**Supplementary Figures 3A,B**).

Transcript Expression, Differentially Expressed Genes, and Differentially Expressed Transcripts

The CPM density distribution of transcript expressions of samples is displayed in **Supplementary Figure 4A**, and the dispersion degree of expression level distribution in a sample is shown with a boxplot in **Supplementary Figure 4B**. The evaluation of the Pearson correlation coefficient among the biological replicates indicated that the *r* range was from 0.682 to 0.963 (**Supplementary Figure 4C**). A clear separation of the control treatment from the SCN4- and SCN5- infected treatments is shown in principal component analysis (PCA) (**Supplementary Figure 4D**).

In total, 2,255 DEGs and 4,167 DETs were identified for all three comparisons; the upregulated and downregulated DEGs or DETs are listed in **Table 2**, and volcano plots for the DEGs and DETs are shown in **Supplementary Figure 5**. There were no overlapping DEGs (**Figure 3F**) or DETs (**Figure 3G**) found between CK-SCN4-up and CK-SCN5-down and vice versa. The annotated numbers of DEGs and DETs with functional annotation database are listed in **Supplementary Table 8**. Only 7 DEGs were found between the SCN4- and SCN5- treatments, including 3 DEGs identified in CK-SCN5, *Glyma.04G061500* (protein serine/threonine activity, GO:0004674), *Glyma.20G205700* (serine-type endopeptidase

inhibitor activity, GO:0004867), and *Glyma.U029900* (function unknown, plasma membrane, GO:0005886), and 4 DEGs were identified in CK-SCN4, *Glyma.03G120700* (calmodulin binding, GO:0005516), *Glyma.10G093900*, *Glyma.11G099300* (a structural constituent of ribosome, GO:0003735), and *Glyma.19G125300* (calmodulin binding, GO:0005516).

Differentially Expressed Genes-Gene Ontology Annotation and Enrichment Analysis

By GO annotation analysis, even though only 7 DEGs were found between the SCN4- and SCN5- infected soybean roots, the SCN4-treated roots displayed 986 DEGs, while the SCN5-treated roots showed 913 DEGs when each was compared with the control. The top GO classifications for both CK-SCN4 and CK-SCN5 treatments in BP, MF, and CC are displayed in **Supplementary Figures 6A,B**, respectively. The top three enriched BPs for both treatments were a response to growth hormone, auxin-activated signaling pathway, and multidimensional cell growth. The top three CCs were plant-type cell wall and intracellular membrane-bound organelle for CK-SCN4, cell part, intracellular part, and intracellular membrane-bound organelle for CK-SCN5. The top two enriched MFs for both CK-SCN4 and CK-SCN5 were quercetin 3-O-glucosyltransferase (GTF, EC:2.4) activity and coniferyl-alcohol GTF activity, and the third MF type was pectinesterase (PE, EC:3.1.1.11) activity for CK-SCN4 and quercetin 4'-O-GTF activity for CK-SCN5. Various numbers of enriched DEGs for BP, MF, and CC were observed between CK-SCN4 and CK-SCN5 (**Figure 4A**). Especially, the numbers of SREs (arrowhead pointing in **Figure 4A**) in CK-SCN4 are greater than those in CK-SCN5, e.g., response to stimulus (SCN4/SCN5, 613/565), signaling (225/181), immune system process (121/84), detoxification (41/32), and binding (582/493) (**Figure 4A**).

Heterodera glycines Race 4 Infection-Induced More Stress Response Elements With More Upregulation Than H. glycines Race 5 Infection

In further analysis with BP annotation of SREs, a greater number of upregulated DEGs than that of downregulated ones was found in both CK-SCN4 and CK-SCN5 (**Figure 4B** and **Supplementary Table 9**). For example, the total number of the up-/down-regulated DEGs added together for each detected SRE was 8,780/6,662 for CK-SCN4 and 7,113/5,932 for CK-SCN5, indicating that more SRE genes were induced in the incompatible reaction than in the compatible reaction (**Figure 4** and **Supplementary Table 9**). The BP response to stimulus included all kinds of biotic and abiotic stresses or stimuli (**Supplementary Table 9**), e.g., pathogens or pests, organic or inorganic chemicals, hormones, temperature, endogenous or external stimulus, defense response, cell death, hypersensitive response, immune response, gene silencing, and MAP kinases (**Figure 4B**).

Since the response to growth hormone was the first top enriched GO-BP for both treatments mentioned above, the response to each hormone was inspected and compared.

TABLE 2 | Number of differentially expressed genes (DEGs) and transcripts (DETs) between un-infected (CK) and *Heterodera glycines*-infected treatments (SCN4 and SCN5), and between SCN5 and SCN4.

	DEG			DET		
	Total number	Up-regulated	Down-regulated	Total number	Up-regulated	Down-regulated
CKvsSCN4	1,533	726	807	2,694	1,121	1,573
CKvsSCN5	1,383	715	668	2,750	1,064	1,686
SCN5vsSCN4	7	6	1	8	7	1

SCN, soybean cyst nematode; vs.: versus.

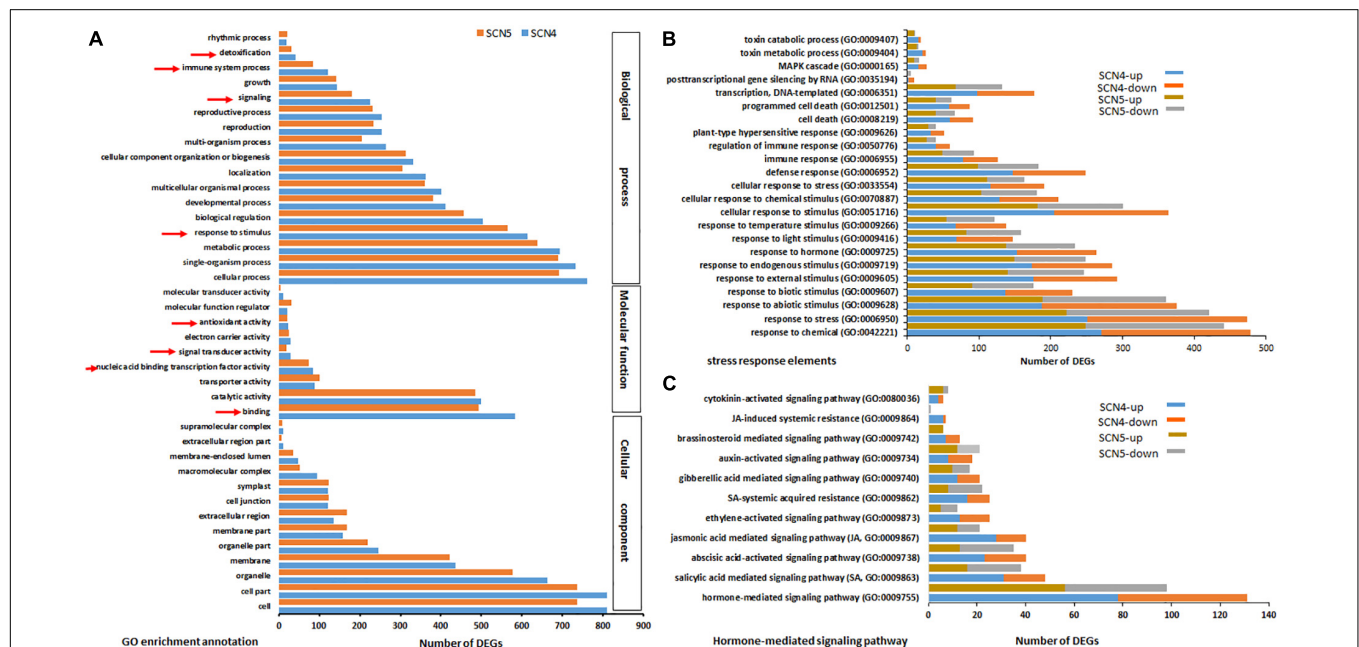


FIGURE 4 | Top GO-enriched annotation of the differentially expressed genes (DEGs) (A) and comparison of DEGs with up-/downregulation associated with stress response elements (B) and hormones (C) in soybean 09-138 infected by soybean cyst nematode race 4 (SCN4) and race 5 (SCN5) when compared with the control (CK). The red arrowheads in A indicate stress response elements.

Different responses to eight types of hormones were identified in both resistant and susceptible reactions compared with the control (Figure 4C and Supplementary Table 9). Although the hormones responded to infection with both nematode races, a greater number of upregulated DEGs than down-regulated ones were found in the resistant reaction to SCN4 in SA- (up/down, 31/17), JA- (28/12), abscisic acid- (ABA, 23/17), ET(13/12), gibberellic acid- (GA, 12/9), brassinosteroid- (BR, 7/6), and cytokinin- (CTK, 4/2) mediated signaling pathways and SA- (16/9) and JA- (6/1) induced systemic resistance except for auxin-activated signaling pathway (IAA, 8/10) (Figure 4C). On the contrary, in the susceptible CK-SCN5 reaction, a lower number of upregulated than downregulated DEGs were found in SA- (16/22), ABA- (13/22), and ET- (5/7) mediated signaling pathways, and SA- (8/14) or JA- (0/1) induced systemic resistance. These results suggested that more SREs with more upregulation were activated by the incompatible reaction with the SCN4 infection when compared with the SCN5 infection.

Heterodera glycines Race 5 Infection Induced More Differentially Expressed Genes With More Upregulated Genes Associated With Cell Wall Modification and Carbohydrate Biological Process Than *H. glycines* Race 4 Infection When Each Was Compared With the Control

Unsurprisingly, 23 (17 up/6 down) and 2 (2 up/0 down) DEGs for response to nematodes and syncytium formation were identified in the resistant soybean reaction to SCN4, respectively, while 34 (24 up/10 down) and 9 (9 up/0 down) DEGs for response to nematodes and syncytium formation were found in the susceptible reaction to SCN5, respectively. SCN syncytium formation requires plant cell wall modification, while cell wall modification involves multidimensional cell growth (SCN4/SCN5 DEGs: 38/48) which was the third BP-enriched element in both treatments. Therefore, all annotations involved in the cell wall were collected together, and a total of 28/26 GOs were annotated for CK-SCN4/CK-SCN5 (Figure 5A). A greater number of DEGs in CK-SCN5 than in CK-SCN4 was detected

in cell wall organization or biogenesis (SCN5/SCN4: 172/159), cell wall modification (51/27), plant-type cell wall modification (27/15), cell wall modification involved in multidimensional cell growth (12/1), plant-type cell wall loosening (11/2), cell wall thickening (11/9), and callose deposition in the cell wall (8/5).

Syncytium formation requires carbohydrates as a nutrition sink for nematode feeding, and sugar transporters are active in syncytia through inter- and intracellular transport processes (Hofmann et al., 2007, 2009; Hofmann and Grundler, 2008). The greater DEG number associated with syncytium formation in CK-SCN5 than in CK-SCN4 denoted the difference in the carbohydrate metabolic process. Thus, the biological processes of carbohydrates and corresponding three major groups (sugars, oligosaccharides, and polysaccharides), including metabolic, biosynthetic, catabolic and transport processes, were examined and compared between CK-SCN4 and CK-SCN5 (Figure 5B). In total, 178, 94, 57, and 13 DEG in CK-SCN4 were involved in carbohydrate metabolic, biosynthetic, catabolic, and transport processes, respectively; 203, 114, 60, and 22 DEGs in CK-SCN5 were associated with the four processes, respectively. For all carbohydrate-associated biological processes, 739 DEGs with 317 up-/422 downregulation in CK-SCN4 and 870 DEGs with 467 up-/403 downregulation in CK-SCN5 were found (Figure 5B). Polysaccharides are major components of carbohydrate metabolic, biosynthetic, and catabolic processes. The lower number of upregulated DEGs (130) than downregulated DEGs (195) for CK-SCN4 was detected in all the groups of carbohydrate metabolic and biosynthetic processes; for CK-SCN5, the greater number of upregulated genes than downregulated genes were only found in polysaccharide metabolic (up/down: 80/51), and biosynthetic (up/down: 55/30) processes and monosaccharide biosynthetic (7/5) process (Figure 5B). Polysaccharides and monosaccharides were a major carbohydrate catabolic process, but polysaccharides displayed more upregulation than downregulation, and monosaccharides showed more downregulation in both CK-SCN4 and CK-SCN5 (Figure 5B). The transport process of carbohydrates demonstrated that oligosaccharides and monosaccharides were major transport formats; only 1 upregulated DEG for polysaccharide transport was detected in CK-SCN5 but not in CK-SCN4. These results suggested that the compatible reaction induced more upregulated expression in the carbohydrate biological process than the incompatible reaction.

Kyoto Encyclopedia of Genes and Genomes Pathway Enrichment Analysis

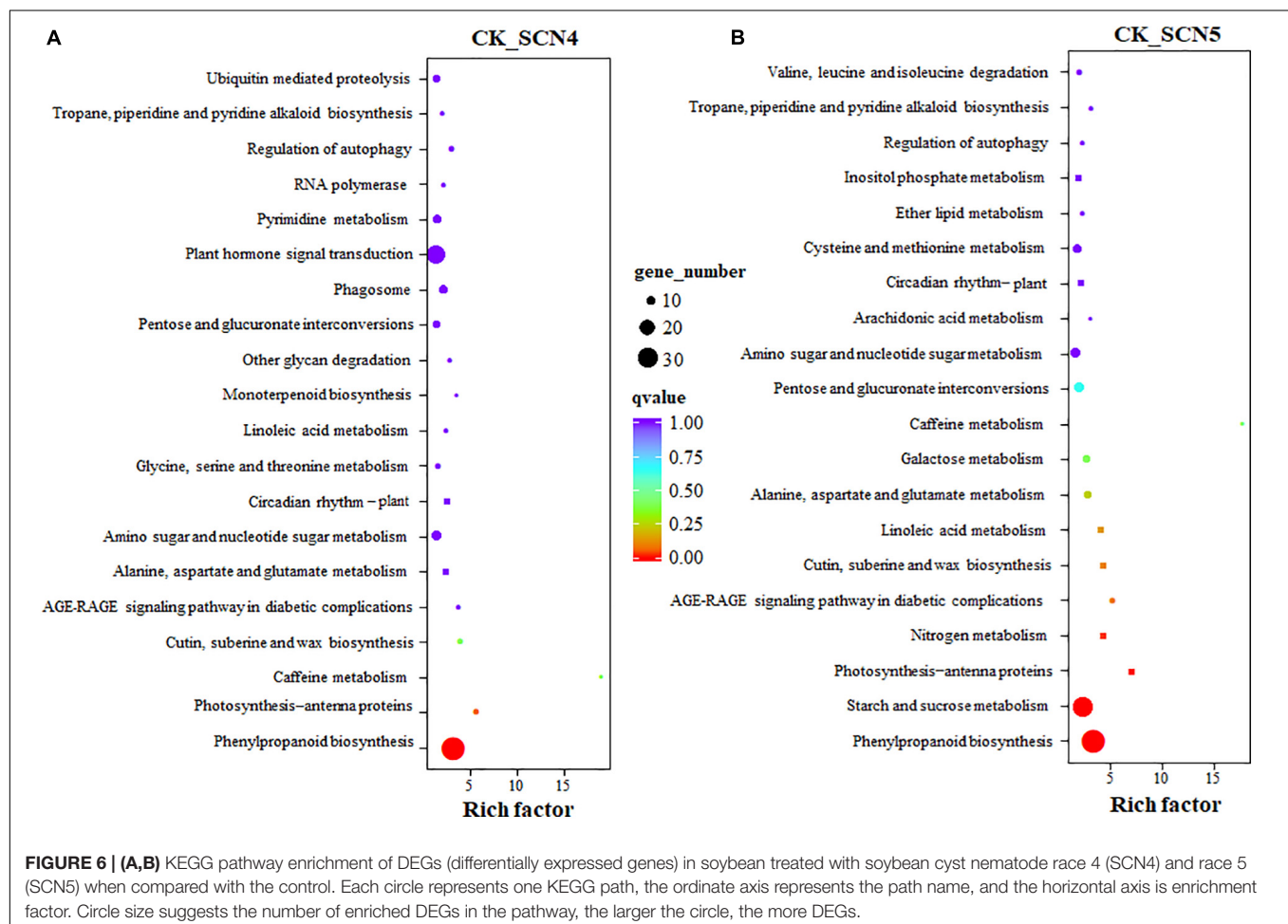
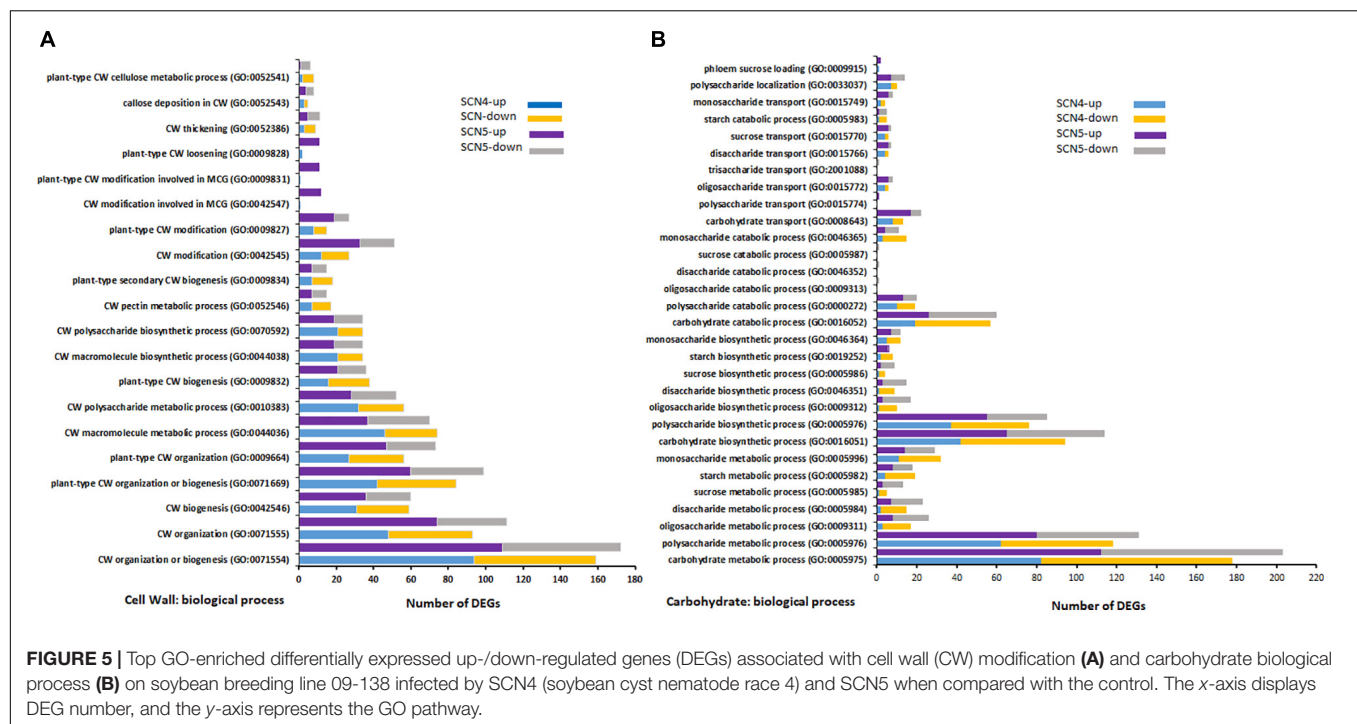
The KEGG annotation showed that 273 and 292 unigenes were involved in 96 pathways in CK-SCN4 and 91 pathways in CK-SCN5, respectively (Supplementary Table 10). Two treatments shared 83 pathways; 13 and 8 unique pathways were detected in CK-SCN4 and CK-SCN5, respectively, e.g., SNARE interactions in vesicular transport (3 DEGs) and oxidative phosphorylation (4 DEGs) only in CK-SCN4 and N-Glycan biosynthesis (4 DEGs) and fatty acid degradation (1 DEG) only in CK-SCN5.

Phenylpropanoid biosynthesis was the top enriched KEGG pathway for both CK-SCN4 (Figure 6A) and CK-SCN5

(Figure 6B). The following top 5 enriched KEGG pathways were photosynthesis-antenna proteins, caffeine metabolism, cutin, suberin and wax biosynthesis, and AGE-RAGE signaling pathway in diabetic complications for CK-SCN4 (Figure 6A), and starch and sucrose metabolism, photosynthesis-antenna proteins, nitrogen metabolism, and AGE-RAGE signaling pathway in diabetic complications for CK-SCN5 (Figure 6B). Among the top 20 enriched KEGG pathways, glycine, serine and threonine metabolism, monoterpenoid biosynthesis, other glycan degradation, phagosome, plant hormone signal transduction, pyrimidine metabolism, and ubiquitin-mediated proteolysis were unique for CK-SCN4; five metabolisms (nitrogen, galactose, cysteine and methionine, ether lipid, and inositol phosphate), and valine, leucine, and isoleucine degradation were unique for CK-SCN5 (Figure 6B). Interestingly, the KEGG pathway circadian rhythm-plant was enriched in the top 8th for CK-SCN4 and in the top 14th for CK-SCN5 (Figures 6A,B).

Phenylpropanoid Biosynthesis Pathway Was Enriched by Both Nematode Infections

Both SCN4 and SCN5 infections induced phenylpropanoid biosynthesis, the top enriched KEGG pathway. Phenylpropanoid compounds are rich sources of metabolism in plants, providing precursors of lignin synthesis (Yao et al., 2021), while lignin is abundant in the cell wall, and it plays vital roles in plant defense in local or systematic required resistance (Dixon et al., 2002). Enzymatic reactions are considered key steps in the biosynthesis of major classes of phenylpropanoid compounds. Thirty-five KEGG orthologies of peroxidase (K00430, EC:1.11.1.7), the richest enzyme annotated in this pathway, were detected in either up- or down-regulation for both CK-SCN4 (up/down: 24/4) and CK-SCN5 (up/down: 23/6); 10 out of 35 *peroxidase* genes were unique for CK-SCN5, and 7 were unique for CK-SCN4 (Figures 7A,B). The peroxidase enzyme plays a role in the final step of the biosynthesis of lignin to produce p-hydroxyphenyl lignin, guaiacyl lignin, 5-hydroxyguaiacyl lignin, and syringyl lignin (Figure 7B). The gene expression levels of *peroxidase 3* (*Glyma.03G208200*, *Glyma.10G022500*) were increased by up to 5.2- to 7.3-fold after nematode infection when compared to the control. The SCN4 infection induced two downregulated DEGs: one gene, *Glyma.15G002600*, encoded enzyme shikimate O-hydroxycinnamoyltransferase (HST, EC:2.3.1.133), which catalyzes the reaction of two substrates, 4-coumaroyl-CoA (4-CCoA) and shikimate, to produce two products, CoA and 4-coumaroylshikimate (CSH); the other DEG, *Glyma.03G070300*, encoded serine carboxypeptidase-like 11 (SCPL11, EC:2.3.1.91), which can form sinapoylcholine (sinapine) and plays vital roles in abiotic and biotic stress responses, and growth and development as well (Xu et al., 2021). The SCN5 infection induced the downregulation of DEG *Glyma.04G227700* encoding caffeic acid 3-O-methyltransferase (COMT)-like (EC:2.1.1.68), a key enzyme regulating lignin synthesis (Trabucco et al., 2013) but not for the SCN4 infection. One gene (*Glyma.15G031300*) encoding the enzyme cyanogenic beta-glucosidase 13-like (EC:3.2.1.21) involved in cell wall lignification increased the expression level 5.87-fold by SCN5 infection (Figure 7A). Both nematode infections activated the upregulation of the



DEG *Glyma.09G201200* encoding enzyme cinnamyl-alcohol dehydrogenase (CAD, EC:1.1.1.195) to catalyze the final step of monolignol biosynthesis, including p-coumaryl alcohol (H), coniferyl alcohol (G), and sinapyl alcohol (S), the main component of lignin (Figures 7A,B). These results indicated the expression variation of peroxidase enzyme, HST, SCPL, and other enzymes associated with lignin synthesis after nematode infection resulted in compatible and incompatible reactions by regulating plant cell wall modification.

Plant Hormone Signal Transduction, Plant–Pathogen Interaction, and MYB Transcription Factors Contributed to Defense Response

A similar number of DEGs were identified in both the plant hormone signal transduction pathway (SCN4/SCN5: 26/25) (Figures 7C,D) and the plant–pathogen interaction pathway (SCN4/SCN5: 12/14) (Figures 7E,F) but with various expressions in levels or gene types. As illustrated in Figures 7C,D, both nematode infections positively or negatively regulated the expression of the hormones (SA, JA, ET, GA, ABA, IAA, and CTK) but with qualitative and quantitative differences in soybean (Figure 7D). For example, the expression level of *PR1* (pathogenesis-related protein 1, *Glyma.15G062400*), which is involved in multiple pathways not only plant hormone signal transduction (SA) but also MAPK signaling and plant–pathogen interaction, was increased by more than 5-fold in soybean by the SCN4 (6.5-fold) and SCN5 (5.2-fold) infections when compared with the control. In the auxin pathway, which is linked to cell enlargement and plant growth, *AUX1* and *AUX/IAA* were upregulated, e.g., auxin transporter-like protein 4 (*Glyma.03G063900*) was positively regulated by both the SCN4 (4.1-fold) and SCN5 (4.5-fold) infections, and auxin-responsive gene *SAUR* (small auxin upregulated RNA) was both up- and downregulated (Figures 7C,D). Two transcription factors PIF3 (phytochrome-interacting factor 3)-like encoded by *Glyma.19G224700* and *Glyma.20G091200*, were identified in branch GA pathway as well as in circadian rhythm plant pathway, with the former being downregulated by both nematode infections and the latter being downregulated by the SCN5 infection. Three genes (*Glyma.09G066500*, *Glyma.19G069200*, and *Glyma.11G018000*) encoding PP2C (protein phosphatase 2C) in the ABA pathway were upregulated; the SCN4 infection induced an approximately 3-fold increase in the expression level of the three genes, and the SCN5 infection increased by 2.2- to 2.8-fold the expression level of the first two genes. JAZ (*jasmonate* zim domain) encoded by *Glyma.17G043700* in JA associated with stress response was detected only in CK-SCN5 with downregulation but not in CK-SCN4. On the contrary, DEGs encoding ETR (ethylene response sensor 2), EBF1 (EIN3-binding F-box protein 1), and ERF1 (ethylene-responsive transcription factor 1) in the ET pathway linked to fruit ripening and senescence were found only in resistant response to SCN4 but not to SCN5 when compared with the control (Figures 7C,D), indicating that the ET pathway played a role in defense response, and that JA expression was inhibited in susceptible response.

In the plant-pathogen interaction pathway, CERK1 (chitin elicitor receptor kinase 1, *Glyma.02G270800*) classified as PRRs in PTI, CDPK (calcium-dependent protein kinase), Rboh (respiratory burst oxidase) triggering reactive oxygen species (ROS) burst correlated to disease HR, CaM/CML (calmodulin/calmodulin-like protein) associated with HR, cell wall reinforcement and stomatal closure, and *PR1* (*Glyma.15G062400*) linked to phytoalexin accumulation and miRNA production, were identified in both CK-SCN4 and CK-SCN5 but with various expression levels; downstream WRKY transcription factor *AtWRKY33* (*GmWRKY15*, *Glyma.02G232600*) and *Pti6* (PR genes transcriptional activator 6, *Glyma.02G236800*) connected to defense-related gene induction were significantly detected only in CK-SCN4 (Figures 7E,F). One gene, *Glyma.01G068000*, encoding HSP90 was detected with downregulation in CK-SCN4, but two other genes, *Glyma.09G131500* (*hsp83*) and *Glyma.16G178800* (*hsp83*), were negatively regulated with significance in CK-SCN5, suggesting that PTI was triggered initially by both nematode infections and that later, ETI was activated by the SCN4 infection.

In the KEGG analysis, 20 MYB or MYB-like transcription factors were detected; of these, 12 were found only in CK-SCN4 and 3 only in CK-SCN5 treatments, denoting that MYB transcription factors might regulate defense response. Three DEGs (*Glyma.03G261800*, *Glyma.16G017400*, and *Glyma.19G260900*) were associated with MYB-related transcription factor LHY involved in plant circadian rhythm as well, were negatively regulated after both SCN4 and SCN5 infections.

Starch and Sucrose Metabolism Pathway Regulated Susceptible Response to *Heterodera glycines* Race 5 and Chitinase I Was Upregulated in Defense Response to *H. glycines* Race 4

The top second KEGG enriched starch and sucrose metabolism pathway in CK-SCN5 treatment had 40 DEGs including 5 sucrose synthases, 11 pectinesterase-like or pectinesterase/pectinesterase inhibitors, 5 beta-glucosidase, and 4 alphas, alpha-trehalose-phosphate synthase [UDP-forming] and others, which are involved in plant growth and cell wall modification or cell wall lignification; these genes are necessary for syncytium formation. Of these, 8/9 and 16/19 up-/downregulated genes were identified for CK-SCN4 and CK-SCN5, respectively (Supplementary Figure 7A), e.g., *Glyma.15G223500* (pectinesterase/pectinesterase inhibitor 47) with a 5-fold increase in expression level after SCN5 infection, while upregulated *Glyma.03G216000* (pectinesterase/pectinesterase inhibitors 20) had 3.3-fold increased expression level in SCN4-infected roots, indicating that these DEGs played roles in the compatible or incompatible reaction. In the amino sugar and nucleotide pathway, 17/3 up-/down regulated DEGs were found, two of which had more than 4-fold significantly increased expression level, *Glyma.02G042500* encoding chitinase class I precursor and *Glyma.18G120700* encoding heveamine-A-like protein in SCN4-infected roots (Supplementary Figure 7B), suggesting that the two genes might contribute to defense response with SCN4 infection.

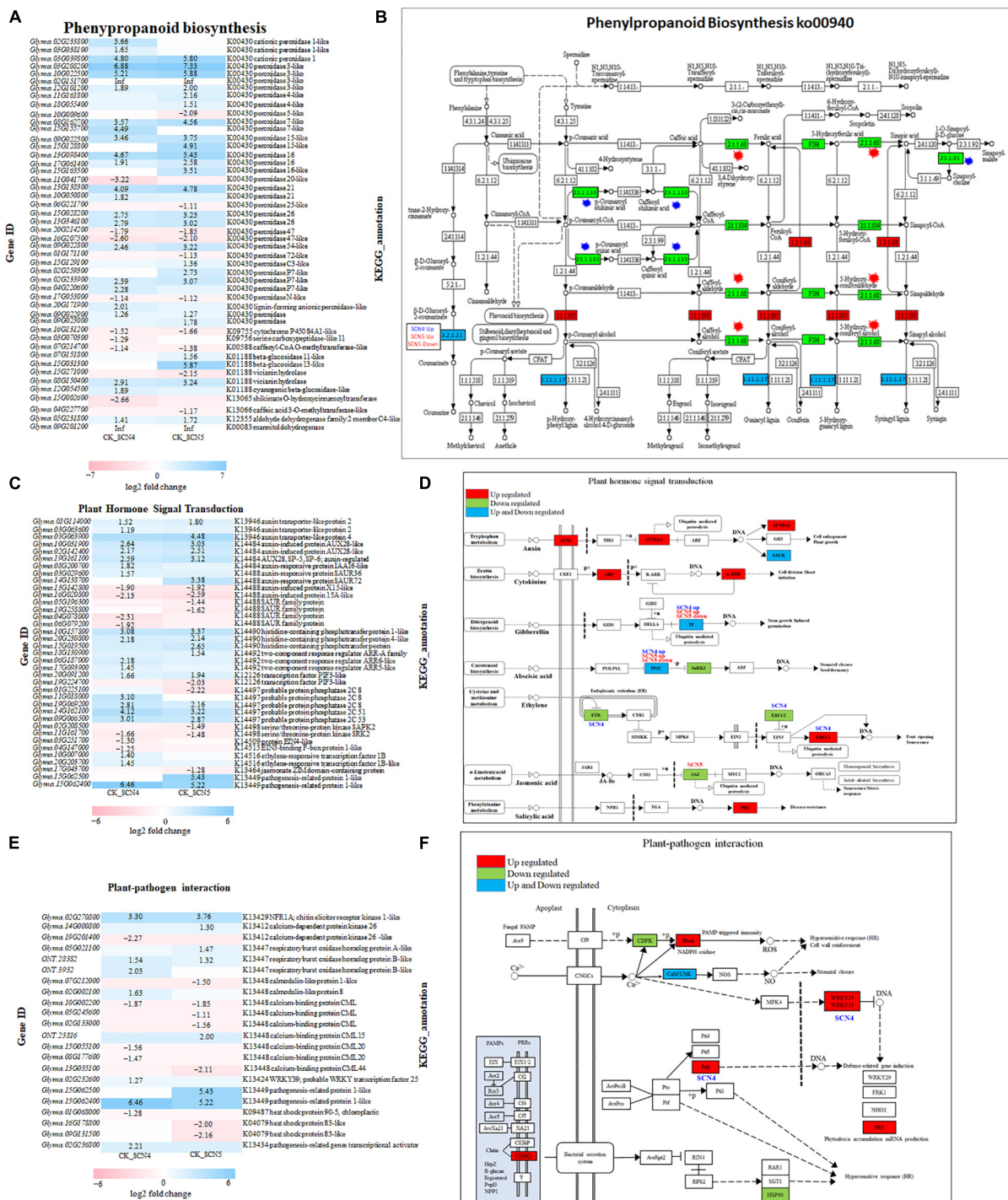


FIGURE 7 | Gene expression and KEGG pathway in phenylpropanoid biosynthesis (<https://www.genome.jp/pathway/map00940>) (**A,B**), plant hormone signal transduction (<https://www.genome.jp/entry/map04075>) (**C,D**) and plant-pathogen interaction (<https://www.genome.jp/entry/map04626>) (**E,F**) in soybean infected with SCN4 (soybean cyst nematode race 4) and SCN5. (**A,C,E**) represent the expression level of differentially expressed genes (DEGs) associated with the corresponding pathway for either CK (control) -SCN4 and/or CK-SCN5; white blank block (0) means no expression variations detected with Fold Change ≥ 2 and False Discovery Rate (FDR) < 0.05 ; the left side of the y-axis is gene ID, and the right side is the corresponding KEGG annotation. (**B,D,F**) are KEGG pathways; the blocks in red, green, and blue colors indicate DEGs in upregulation, downregulation, or both ways for both CK-SCN4 and CK-SCN5, respectively, except for the special note for DEGs only found in either SCN4/SCN5 or either up-/downregulation; in panel (**B**), blue star represents the enzyme DEG found only in CK-SCN4, and the red star represents the enzyme DEG found only in CK-SCN5.

Protein–Protein Interaction Analysis Revealed Valine-Glutamine-WRKY Interactions and Cytokinin Two-Component Response Regulator ARR-Glutaredoxins-BolA Interactions Might Be Involved in Defense Response

The PPI analysis among all the 2,255 DEGs indicated a ratio of 7,063/2,632 (2.7 ×) of protein-protein networks for CK-SCN4/CK-SCN5, including 866/194 activation (4.5 ×), 2715/1381 (2 ×) binding, 781/280 (2.8 ×) catalysis, 401/104 (3.9 ×) expression, 462/96 (4.8 ×) inhibition, 874/166 (5.3 ×) ptmod (post translational modification), and 964/411 (2.3 ×) reaction, indicating soybean resistant response to the SCN4 infection activated more protein–protein interactions than a susceptible response to the SCN5 infection, especially in post translational modification, inhibition, and activation.

Two genes, *Glyma.03G120700* (*GmVQ5*) and *Glyma.19G125300* (*GmVQ70*), were identified as significant DEGs with a 2.7–3.2-fold increase in CK-SCN4 than in CK-SCN5, encode calmodulin binding protein (CaMBP) (GO:0005516) carrying conserved VQ (Valine-Glutamine) motif (FxxhVQxhTG), which is capable of interacting with the transcription factor WRKY DNA-binding domain to play vital roles in stress response (Wang et al., 2014). Thus, the interactions between VQ and WRKY proteins encoded by DEGs were explored based on the Pfam domain; 6 (*GmVQ5*, *GmVQ8*, *GmVQ24*, *GmVQ43*, *GmVQ44*, and *GmVQ70*) out of 26 VQs and 19 out of 74 WRKYs annotated were detected in either CK-SCN4, CK-SCN5, or SCN5-SCN4 (**Figure 8A**). Interestingly, *GmWRKY15*, encoded by the upregulated *Glyma.02G232600*, homologous with *Arabidopsis* AtWRKY33 (WRKYGQK/WRKYGQK) in the plant–pathogen interaction pathway, was identified in CK-SCN4, which could interact with *GmVQ24* (AtVQ21, *Glyma.06G124400*), and both were linked to *GmVQ44* (AtVQ25, calcium-binding protein, *Glyma.09G111800*) with downregulation, which could bind with the two homologous CaM-binding proteins *GmVQ5* and *GmVQ70* (**Figure 8B**). VQ and WRKY numbers were designed according to Yu et al. (2016), Wang et al. (2019), respectively.

AtWRKY33 can interact with a VQ protein called mitogen-activated protein (MAP) kinase substrate1 (MSK1, AtVQ21), a substrate of MPK4 (MAP kinase 4), which can be activated in the presence of pathogen attack or flagellin, and then AtWRKY33 is released to induce the expression of phytoalexin deficient 3 (PAD3) in the nucleus, and thereby to increase defense response (Andreasson et al., 2005; Cheng et al., 2012). In addition, MPK3 and MPK6 can also interact with AtWRKY33 together, leading to an increase in phytoalexin-related gene expression (Alves et al., 2014). The overexpression of AtWRKY33 resulting in decreased susceptibility to beet cyst nematode (*Heterodera schachtii*) (Ali et al., 2013) supports that AtWRKY33 may be involved in soybean cyst nematode resistance. To find the kinase of the substrate *GmVQ24* (MSK1) in the annotation data, all expressed MAP kinases were searched

and no homologous AtMPK4 was found; interestingly, only two homologous MPK3s encoded by *Glyma.U021800* and *Glyma.12G073000*, each with a 1-fold increase in expression level in CK-SCN4, were identified to interact with *GmWRKY15* and *GmVQ24*. Further searching of other DEGs encoding MAPKs exhibited that two genes encoding MAP kinase kinase 2 (MPKK2) had a 1.5- to 1.8-fold reduction in the expression level only in CK-SCN4 (**Figure 8B**). Two more MPKKs encoded by *Glyma.17G173000* and *Glyma.17G177900*, which could bind to MPK3 and MPKK2, were found with a 1.2-fold increase of gene expression in CK-SCN4 (data not shown). All these results demonstrated that MPKKK/MPKK/MPK/WRKY-VQ-CaMBP-VQ might work together in soybean to defend against SCN4 infection. Based on these data, a model with a VQ–WRKY interaction in response to SCN4 infection was established with two possibilities, one is *GmWRKY33* release by MPK3 activation or any other kinase to induce phytoalexin production, which activates plant defense as described in *Arabidopsis* (Alves et al., 2014); the other is *GmVQ44* binding to *GmVQ24* (MSK1) to suppress *GmVQ44* expression, which results in high expression of CaMBP *GmVQ5/70* to induce SCN resistance (**Figure 8C**).

Another interesting protein–protein interaction was detected in the plant hormone signal branch cytokinin pathway with two upregulated CK-SCN4 DEG genes (**Figure 8D**), *Glyma.06G187000* (*ARR6*, Log₂FC = 2.2) and *Glyma.17G093900* (*ARR5*, Log₂FC = 1.5), which are classified as two-component response regulator ARR-A (type A *Arabidopsis* response regulator) family composed of an inner membrane-spanning histidine kinase and a cytoplasmic response regulator to allow organisms to sense and respond to changes in environmental stimuli (Stock et al., 2000). In CK-SCN4, the ARR5-like and ARR6-like proteins were capable of interacting with a mitochondrial chaperone, monothiol glutaredoxin-S15 (GRXS15), and BolA2. GRXS15 encoded by the downregulated DEG *Glyma.08G209900* (Log₂FC = -1.8), and BolA2 encoded by a downregulated DEG *Glyma.05G188300* (Log₂FC = -1.6) could connect with the transcription factor GATA zinc finger (*Glyma.06G086400*, Log₂FC = 1.7) (**Figure 8C**). Mitochondria GRXS15 was reported to play key roles in iron-sulfur protein maturation in the plant (Moseler et al., 2015) and a transcription regulator BolA2 containing a helix-turn-helix (HTH) motif for nucleic acid binding (Kasai et al., 2004). AtBolA3–GRXS17 interaction plays a key role in suppressing abiotic tolerance (Cheng et al., 2011; Qin et al., 2015). In a similar manner, inhibition of the gene expression of both BolA2 and GRXS15 with SCN4 infection resulting in resistance denoted that BolA2–GRXS15 interaction may negatively regulate soybean resistance to SCN4. The function of BolA2–GRXS15 interaction has not been reported yet, but it is known that BolA2 is nucleon-cytoplasmic and interacts with GRXS15 (Couturier et al., 2014). Furthermore, BolA2 was linked to the lncRNA detected above. Therefore, the interactions of ARR5/6–GRXS15–BolA2–GATA might indicate a new complex interaction partaking in plant defense in response to SCN4 infection.

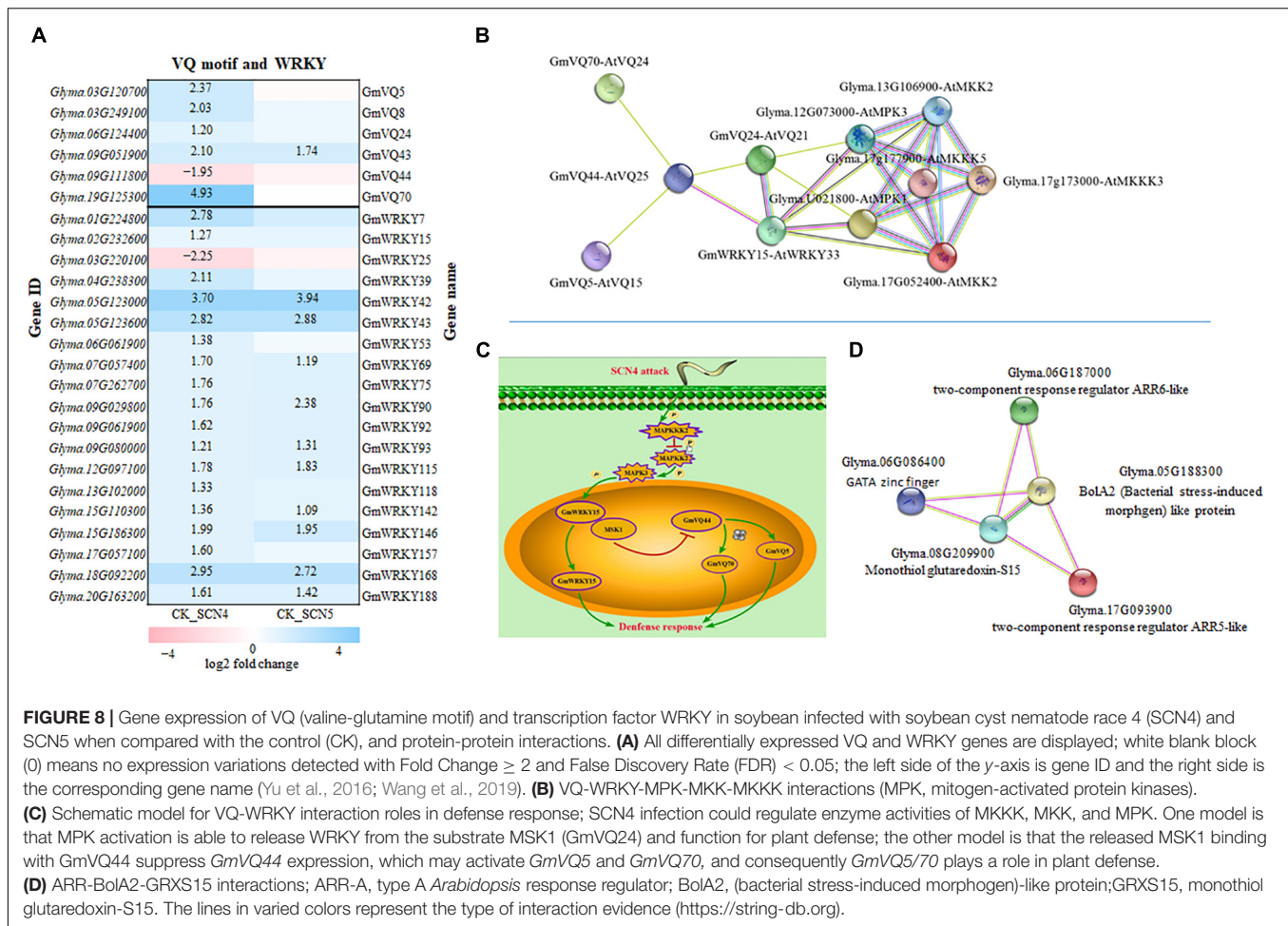


FIGURE 8 | Gene expression of VQ (valine-glutamine motif) and transcription factor WRKY in soybean infected with soybean cyst nematode race 4 (SCN4) and SCN5 when compared with the control (CK), and protein-protein interactions. **(A)** All differentially expressed VQ and WRKY genes are displayed; white blank block (0) means no expression variations detected with Fold Change ≥ 2 and False Discovery Rate (FDR) < 0.05 ; the left side of the y-axis is gene ID and the right side is the corresponding gene name (Yu et al., 2016; Wang et al., 2019). **(B)** VQ-WRKY-MPK-MKK-MKKK interactions (MPK, mitogen-activated protein kinases). **(C)** Schematic model for VQ-WRKY interaction roles in defense response; SCN4 infection could regulate enzyme activities of MKKK, MKK, and MPK. One model is that MPK activation is able to release WRKY from the substrate MSK1 (GmVQ24) and function for plant defense; the other model is that the released MSK1 binding with GmVQ44 suppress GmVQ44 expression, which may activate GmVQ5 and GmVQ70, and consequently GmVQ5/70 plays a role in plant defense. **(D)** ARR-BolA2-GRXS15 interactions; ARR-A, type A *Arabidopsis* response regulator; BolA2, (bacterial stress-induced morphogen)-like protein; GRXS15, monothiol glutaredoxin-S15. The lines in varied colors represent the type of interaction evidence (<https://string-db.org>).

Validation of Differentially Expressed Genes by QRT-PCR Assay

The expression levels of 23 DEGs obtained from full-length transcriptome sequencing were compared with those obtained from qRT-PCR. The expression correlation (R^2) between full-length-seq and qRT-PCR was up to 0.6958 (Figure 9A), and the comparison of relative expression levels between SCN4 and SCN5 demonstrated that the qRT-PCR data almost matched with the full-length-seq data (Figure 9B). The relative expression levels of four WRKY and three VQ genes are also displayed in Figure 9C.

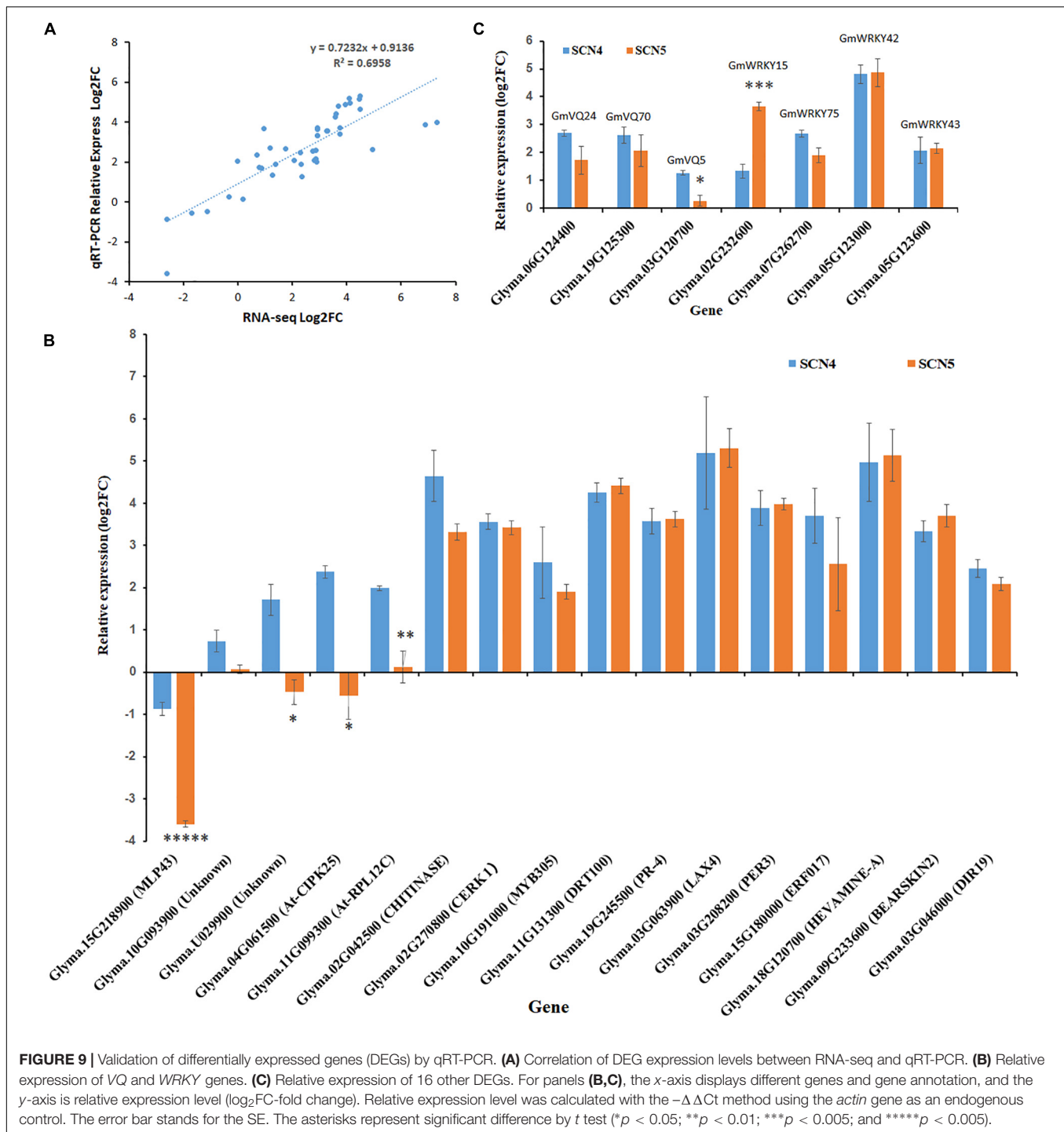
DISCUSSION

Full-Length Transcriptome Analysis: A Powerful Tool to Analyze Transcriptional and Post-transcriptional Regulation

In this study, full-length transcriptome analysis on soybean was first conducted to compare responses of the same soybean genotype incompatible and compatible to different *H. glycines* races. Obviously, the advantages of full-length transcriptome sequencing include high-cost performance, high throughput, no

GC specificity, and base bias, long sequencing reads, accurate quantification at the transcriptome level, accurate identification of structure characteristics (e.g., AS, fusion gene, APA) without breaking sequences or gene structures, and DEG/DET analysis at once. Compared to NGS, ONT needs less read numbers to cover the same amount of transcripts. For example, an average of $3\text{--}5 \times$ greater number of clean reads was obtained by NGS (Zhang et al., 2017; Neupane et al., 2019; Miraeiz et al., 2020) when compared with the average number of clear reads in this study. In addition, the identified novel genes can generate new AS events.

The change in numbers of AS events at 5' splicing sites or at 3' splicing sites after both nematode infections and various transcripts among the three treatments indicated that the nematode infections did trigger differential AS events to generate multiple transcripts that might result in plant defense or susceptibility. A few cases have been reported that AS regulates plant stress responses and/or adaptations (Bedre et al., 2019; John et al., 2021; Martín et al., 2021). For instance, the tobacco gene *N* resistant to tobacco mosaic virus (TMV) can be alternatively spliced to produce two transcripts that are required to contribute to complete resistance to TMV (Dinesh-Kumar and Baker, 2000). AS regulates ABA and light signaling pathways to coordinate plant growth and stress response (Thiruppathi, 2020).



Interestingly, we found that more abundant APA events per gene were identified with > 5 poly A sites (23.8%) than those in 1–5 poly A sites in soybean, while 2 poly A sites were found as the most abundant APA event distribution in sorghum (Chakrabarti et al., 2020), and 1 poly A site for the tree species *Liriodendron chinense* in the magnolia family (Tu et al., 2021). The poly A site distribution in plants can be shifted by abiotic stresses (Yan et al., 2021), and novel stress-specific cis-elements in intronic

poly A sites may contribute to abiotic stresses (Chakrabarti et al., 2020). In this study, nematode infection increased the > 5-poly A site distribution than control, especially, the > 10 poly A sites in the incompatible reaction further confirmed that APA may be involved in nematode stress response. Generally, stress-responsive APA plays a key role in regulating abiotic stress and biotic stress through crucial stress-responsive genes and pathways (Ye et al., 2019; Chakrabarti et al., 2020;

Yan et al., 2021). For instance, Ye et al. (2019) found that APA in rice responds to heat stress tolerance as a negative regulator, to Cd stress by regulating DNA repair and cell wall formation, and to disease stress (bacterial blight, rice stripe virus, and rice blast) by regulating chlorophyll metabolism. Comparisons of APA sites linked with annotated genes or pathways between control and nematode-infected treatments will uncover the APA function in soybean associated with nematode resistance or susceptibility. Furthermore, the difference in the fusion transcripts and lncRNAs among the treatments demonstrates that full-length transcriptome sequencing is a powerful tool to analyze post-transcriptional modification.

Plant Hormone Signal Transduction and Plant–Pathogen Interaction Implicating Plant Defense Response

Both the GO and KEGG enrichment analyses consistently confirmed that stress response elements and associated pathways (plant hormone signal transduction and plant–pathogen interaction) contribute to plant innate immune response to SCN. Surprisingly, when all the DEGs in these groups/pathways identified in this study were compared with those listed by Zhang et al. (2017), Miraeiz et al. (2020), only 1–3 overlapping genes were found; the only DEG, *Glyma.11G207000*, encoding leucine-rich repeat-containing protein was found in all three studies. The difference might be caused by various SCN–soybean interaction systems and be affected by inoculation time, genotypes, nematode types, and possible nematode inoculation density. For example, a transcriptome analysis was conducted as early as 8 h post inoculation with HG type 0 on soybean Peking, *G. soja* PI 468916, Fayette, and Williams 82 by Miraeiz et al. (2020), and later sedentary phase at 3, 5 and 8 days (pooled samples) with HG type 2.5.7 on two *G. soja* genotypes by Zhang et al. (2017), and 8 days with HG Type 2.5.7 and HG Type 1.2.3.5.6.7 on the same genotype 09-138 in this study. Most transcriptome profiling studies are examined within 2–10 days since SCN feeding site establishment generally required 48 h after inoculation, and syncytium formation and syncytium collapse happens within 2–10 days after inoculation in resistance response (Klink et al., 2007; Kandoth et al., 2011).

Although uniquely differential expression genes to SCN4 or SCN5 infection were identified on 09-138, most DEGs were expressed in both resistant and susceptible reactions with only a small difference (Figure 7), indicating qualitative and quantitative traits. For example, in the plant–pathogen interaction pathway, upstream PPR-CERK1 and downstream Ca^{2+} -dependent signal components CDPK, Rboh (triggering reactive oxygen burst, ROS), CaM/CML, HSP, and PR1 were all identified in both compatible and incompatible roots, and downstream WRKYs and Pti6 with upregulation were only detected in the incompatible response (Figures 7E,F), while these components are involved in PTI and/or ETI activity. In the plant-hormone pathway, all phytohormones except for brassinosteroid were activated with 1–4 key DEGs, while the phytohormone networks of the JA, ET, and SA signaling pathways are required for PTI and ETI as well (Cui et al., 2015).

Naveed et al. (2020) found that both resistant and susceptible plants showed almost identical transcriptome responses, but that the resistant plants achieved high-amplitude transcriptional reprogramming several hours earlier than the susceptible plants through a defense phytohormone signaling network. Here, we only detected one time point in the later infection stage, and a time series examination will reveal more about how the pathways work together to defend against nematode attacks. The PTI-ETI continuum concept with crosstalk between PTI-ETI signal components effectively activating plant immune responses has the increasing attention of biologists (Mine et al., 2018; Naveed et al., 2020; Yuan et al., 2021), and this concept may also be applicable for SCN–soybean interaction based on the identified signal components in these pathways. In addition, surprisingly, even though 09-138 contained the Peking-*rhg1a* locus, there were not any DEGs found in that region, suggesting an alternative resistance mechanism in 09-138.

Transcription Factor Functional Protein–Protein Interactions in Soybean Defense Response

Transcription factors as transcriptional regulators function by binding to the promoter region of target genes and regulating plant response to environmental stress, e.g., altering the expression of cascades of defense genes (Chen et al., 2002; Alves et al., 2014). The novel transcripts identified exhibited more than 5,000 TFs, with the top 5 being WRKY, AP2/ERF-ERF, NAC, GRAS, and bHLH. WRKY TFs, as the largest family of transcriptional regulators, were consistently identified in almost all transcriptome analyses of SCN infection (Ithal et al., 2007; Kandoth et al., 2011; Mazarei et al., 2011; Wan et al., 2015; Zhang et al., 2017; Song et al., 2019; Jiang et al., 2020; Miraeiz et al., 2020). The WRKY family, associated with stress responses in soybean, has been examined, e.g., in response to soybean rust disease (*Phakopsora pachyrhizi*) (Bencke-Malato et al., 2014), salt stress (Yu et al., 2016), dehydration and salt stress (Song et al., 2016), and soybean cyst nematode (Yang et al., 2017). However, 19 WRKY-DEGs to SCN4 or SCN5 identified in this study were not found in previous studies for SCN infection (Yang et al., 2017; Zhang et al., 2017; Miraeiz et al., 2020), indicating that these identified WRKY factors may be specific to the line 09-138 or the nematode races used in this study.

WRKY factors in plants are divided into five groups (I, IIa + IIb, IIc, IID + IIE, and III), and the interaction between the WRKY domain and its partner domain (e.g., VQ) is involved in signaling, transcription, and other important biological processes (Chi et al., 2013; Alves et al., 2014). The WRKY-VQ interaction has been well-studied in *Arabidopsis*. As mentioned above, AtWRKY33 (group I)/MSK1 interaction activates plant defense gene expression (Andreasson et al., 2005; Qiu et al., 2008; Alves et al., 2014). AtWRKY25, a homolog of AtWRKY33, is also able to interact with MSK1 and MPK4 in the absence of a pathogen (Cheng et al., 2012). Additionally, AtWRKY33 is able to interact with two other VQ proteins, AtSIB1 (sigma factor binding protein 1, AtVQ23) and AtSIB2 (AtVQ16), in the nucleus to activate

resistance to a necrotrophic pathogen, *Botrytis cinerea* (Lai et al., 2011). In this study, the lack of homologous MPK4 indicates a possible alternate interaction mechanism participating in nematode defense response. In the soybean–pest interaction system, only one GmVQ58 was identified as a negative regulator for soybean resistance to the common cutworm (*Spodoptera litura* Fabricius), and GmVQ58 could interact with GmWRKY32 (Li X. et al., 2020).

The CaM family is composed of ubiquitous Ca^{2+} -binding proteins or calcium sensor proteins, which can play a key role in cellular signaling cascades coupling various environmental stimuli (Zeng et al., 2015). For instance, AtCaMBP25/AtVQ15 was found as a negative effector regulating osmotic stress tolerance during seed germination and seedling growth (Perruc et al., 2004). CaM-mediating signaling can regulate ROS homeostasis directly and indirectly (Zeng et al., 2015). Here, we found DEGs *CaM/CML* in the plant–pathogen pathway and two *CaMBP*GmVQ5/VQ70s, which directly or indirectly interact with other VQs and WRKYs to form complex GmVQ5/70–GmVQ44–GmWRKY15–GmVQ24 (MSK1) (Figure 8B). Only the incompatible reaction could activate GmVQ5/70, GmWRKY15, and GmVQ24, and suppress the connector GmVQ44, indicating that VQ–WRKY interactions partake in plant defense response to SCN4. Additionally, the high expression level of *GmWRKY15* increasing SCN resistance is consistent with that the homologous *AtWRKY33* overexpression decreasing *H. schachtii* susceptibility (Ali et al., 2013).

Top Enriched Phenylpropanoid Biosynthesis Pathway Involved in Both Resistant and Susceptible Response

The phenylpropanoid pathway has been considered a ubiquitous defense response against pathogens including nematodes, and this pathway is always enriched in SCN–soybean interaction (Edens et al., 1995; Dixon et al., 2002; Zhang et al., 2017; Li et al., 2018; Singh et al., 2019; Miraeiz et al., 2020). Approximately 75% of DEGs in this pathway were peroxidase, and more than 80% of the enriched peroxidase genes could be induced after nematode infection either in susceptible or in resistant response, which is matched with previous reports (Miraeiz et al., 2020), suggesting that peroxidase plays a central role in response to nematode attack. Peroxidase contributes to plant defense by modifying the cell wall composed of lignin, suberin, feruloylated polysaccharides, and HPRG (extensins), enhancing ROS production and phytoalexin productions (Pandey et al., 2017). The key enzymes linked to lignin biosynthesis and uniquely expressed in the incompatible (e.g., HST, *Glyma.15G002600*) or compatible (e.g., beta-glucosidase, *Glyma.15G031300*) reaction will be potentially targeted genes for further studies to elucidate plant defense or pathogenesis.

Carbohydrate Biological Process and Cell Wall Modification Roles in Soybean Cyst Nematode Susceptibility

SCN5 infection-induced more DEGs with more upregulated genes in cell wall modification and carbohydrate metabolic

process than SCN4 infection but with some quantitatively overlapping DEGs, indicating unique and accumulative DEGs in carbohydrate biological process including metabolism, biosynthesis, catalyst and transport work together to make SCN susceptibility or resistance. Miraeiz et al. (2020) also reported that susceptible W82 showed less responsive genes, more downregulated genes in carbohydrate metabolism and transport proteins, and some overlapping genes with the resistant genotype 8 h post inoculation. Nevertheless, again, there are less common DEGs found between 8 h post inoculation (Miraeiz et al., 2020) and 8 days (this study), a beta-glucosidase was downregulated at 8 h but upregulated at 8 days in this study, suggesting spatially and temporally differential expression. In the starch and sucrose metabolism pathway, sucrose synthase, hexokinase, beta-amylase, and polygalacturonase were suppressed more, and beta-glucosidase, beta-fructofuranosidase, and galacturonosyltransferase were induced more in the compatible interaction than in the incompatible interaction, proving the complexity for nematode pathogenicity. Hofmann et al. (2007) demonstrated that sucrose supply to *H. schachtii*-induced syncytia depends on the apoplastic pathway in the early stage during syncytium formation and on the symplasmic pathway in the later stage when syncytia are linked to the phloem. The change of a series of enzymes during a metabolic process may explain various enzyme expression patterns in different SCN–soybean interaction systems both spatially and temporally.

CONCLUSION

In conclusion, this study represents for the first time a full-length transcriptome sequencing comparison between compatible and incompatible reactions on the same soybean genome to *H. glycine*. Stress response elements, plant pathogen interaction pathway, plant hormone signaling transduction pathway, and transcription factors contributed to plant immune response. The related genes associated with cell wall modification and carbohydrate metabolism played critical roles in nematode susceptibility. The phenylpropanoid biosynthesis pathway was enriched by two nematode infections. For the first time, a model of WRKY–VQ interaction resulting in plant defense response to nematode infection in an incompatible reaction was established. The identified AS events, APA, and lncRNA will provide insights into the function of post transcriptional modification during plant–nematode interaction. The knowledge of SCN–soybean interaction will aid us to understand the evolution of resistance and susceptibility, and further functional studies will help to explore new control strategies against nematodes.

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: NCBI; PRJNA803218.

AUTHOR CONTRIBUTIONS

CW and CL conceived and designed the study, and wrote, reviewed, and edited the manuscript. MH, YJ, RQ, DJ, and DC performed the laboratory work and conducted the data analysis. MH and YJ wrote the original draft. ZT was a curator of plant material. CW and MH validated the data and analysis. All authors read and approved the final manuscript.

FUNDING

This study was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA24010307) and the National Natural Science Foundation of China (31772139) to CW.

ACKNOWLEDGMENTS

We thank Philip A Roberts (University of California, Riverside, CA, United States) for the helpful comments on the manuscript.

REFERENCES

- Acharya, K., Tande, C., and Byamukama, E. (2016). Determination of *Heterodera glycines* virulence phenotypes occurring in South Dakota. *Plant Dis.* 100, 2281–2286. doi: 10.1094/PDIS-04-16-0572-RE
- Ali, M. A., Abbas, A., Kreil, D. P., and Bohlmann, H. (2013). Overexpression of the transcription factor RAP2.6 leads to enhanced callose deposition in syncytia and enhanced resistance against the beet cyst nematode *Heterodera schachtii* in *Arabidopsis* roots. *BMC Plant Biol.* 13:47. doi: 10.1186/1471-2229-13-47
- Alves, M. S., Dadalto, S. P., Gonçalves, A. B., de Souza, G. B., Barros, V. A., and Fietto, L. G. (2014). Transcription factor functional protein-protein interactions. *Proteomes* 2, 85–106. doi: 10.3390/proteomes2010085
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11:R106. doi: 10.1186/gb-2010-11-10-r106
- Andreasson, E., Jenkins, T., Brodersen, P., Thorgrimsen, S., Petersen, N. H., Zhu, S., et al. (2005). The MAP kinase substrate MSK1 is a regulator of plant defense responses. *EMBO J.* 24, 2579–2589. doi: 10.1038/sj.emboj.7600737
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119. doi: 10.1093/nar/gkh131
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Bailey, T. L., Williams, N., Misleh, C., and Li, W. W. (2006). MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34, 369–373. doi: 10.1093/nar/gkl198
- Bandara, A. Y., Weerasooriya, D. K., Bradley, C. A., Allen, T. W., and Esker, P. D. (2020). Dissecting the economic impact of soybean diseases in the United States over two decades. *PLoS One* 15:e0231141. doi: 10.1371/journal.pone.0231141
- Bedre, R., Irigoyen, S., Schaker, P. D. C., Monteiro-Vitorello, C. B., Da Silva, J. A., and Mandadi, K. K. (2019). Genome-wide alternative splicing landscapes modulated by biotrophic sugarcane smut pathogen. *Sci. Rep.* 9:8876. doi: 10.1038/s41598-019-45184-1
- Bencke-Malato, M., Cabreira, C., Wiebke-Strohm, B., Bucker-Neto, L., Mancini, E., Osorio, M. B., et al. (2014). Genome-wide annotation of the soybean WRKY family and functional characterization of genes involved in response to *Phakopsora pachyrhizi* infection. *BMC Plant Biol.* 14:236. doi: 10.1186/s12870-014-0236-0

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.866322/full#supplementary-material>

Supplementary Table S1 | Primer sequences for qRT-PCR.

Supplementary Table S2 | Statistics of clean data.

Supplementary Table S3 | All fusion gene list for different samples.

Supplementary Table S4 | Summary of SSR analysis.

Supplementary Table S5 | All identified transcription factors.

Supplementary Table S6 | Novel isoform functional annotation.

Supplementary Table S7 | Novel gene functional annotation.

Supplementary Table S8 | Number of annotated DEGs and DETs.

Supplementary Table S9 | Top DEG-GO annotation of stress response element comparison between CK vs SCN4 and CK vs SCN5.

Supplementary Table S10 | Top DEG-KEGG pathway comparison between CK vs SCN4 and CK vs SCN5.

- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Series B. Stat. Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Budak, H., Kaya, S. B., and Cagirci, H. B. (2020). Long non-coding RNA in plants in the era of reference sequences. *Front. Plant Sci.* 11:276. doi: 10.3389/fpls.2020.00276
- Byrd, D. W., Kirkpatrick, T., and Barker, K. R. (1983). An improved technique for clearing and staining plant tissues for detection of nematodes. *J. Nematol.* 15, 142–143. doi: 10.1007/11527503_12
- Cao, J., Ye, C., Hao, G., Dabney-Smith, C., Hunt, A. G., and Li, Q. Q. (2019). Root hair single cell type specific profiles of gene expression and alternative polyadenylation under cadmium stress. *Front. Plant Sci.* 10:589. doi: 10.3389/fpls.2019.00589
- Chakrabarti, M., de Lorenzo, L., Abdel-Ghany, S. E., Reddy, A., and Hunt, A. G. (2020). Wide-ranging transcriptome remodelling mediated by alternative polyadenylation in response to abiotic stresses in Sorghum. *Plant J.* 102, 916–930. doi: 10.1111/tj.14671
- Chen, W., Provart, N. J., Glazebrook, J., Katagiri, F., Chang, H. S., Eulgem, T., et al. (2002). Expression profile matrix of *Arabidopsis* transcription factor genes suggests their putative functions in response to environmental stresses. *Plant Cell* 14, 559–574. doi: 10.1105/tpc.010410
- Cheng, N. H., Liu, J. Z., Liu, X., Wu, Q., Thompson, S. M., Lin, J., et al. (2011). *Arabidopsis* monothiol glutaredoxin, AtGRXS17, is critical for temperature-dependent postembryonic growth and development via modulating auxin response. *J. Biol. Chem.* 286, 20398–20406. doi: 10.1074/jbc.M110.201707
- Cheng, Y., Zhou, Y., Yang, Y., Chi, Y. J., Zhou, J., Chen, J. Y., et al. (2012). Structural and functional analysis of VQ motif-containing proteins in *Arabidopsis* as interacting proteins of WRKY transcription factors. *Plant Physiol.* 159, 810–825. doi: 10.1104/pp.112.196816
- Chi, Y., Yang, Y., Zhou, Y., Zhou, J., Fan, B., Yu, J. Q., et al. (2013). Protein-protein interactions in the regulation of WRKY transcription factors. *Mol. Plant* 6, 287–300. doi: 10.1093/mp/sst026
- Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. (2017). An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Res.* 27, 885–896. doi: 10.1101/gr.217117.116
- Cook, D. E., Bayless, A. M., Wang, K., Guo, X., Song, Q., Jiang, J., et al. (2014). Distinct copy number, coding sequence, and locus methylation patterns

- underlie *Rhg1*-mediated soybean resistance to soybean cyst nematode. *Plant Physiol.* 165, 630–647. doi: 10.1104/pp.114.235952
- Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., et al. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in soybean. *Science* 338, 1206–1209. doi: 10.1126/science.1228746
- Couturier, J., Wu, H. C., Dhalleine, T., Pégeot, H., Sudre, D., Gualberto, J. M., et al. (2014). Monothiol glutaredoxin-BolA interactions: redox control of *Arabidopsis thaliana* BolA2 and SufE1. *Mol. Plant* 7, 187–205. doi: 10.1093/mp/sst156
- Cui, H., Tsuda, K., and Parker, J. E. (2015). Effector-triggered immunity: from pathogen perception to robust defense. *Annu. Rev. Plant Biol.* 66, 487–511. doi: 10.1146/annurev-arplant-050213-040012
- Cui, J., Shen, N., Lu, Z., Xu, G., Wang, Y., and Jin, B. (2020). Analysis and comprehensive comparison of PacBio and nanopore-based RNA sequencing of the *Arabidopsis* transcriptome. *Plant Methods* 16:85. doi: 10.1186/s13007-020-00629-x
- Dangl, J. L., and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* 411, 826–833. doi: 10.1038/35081161
- Deepak, S., Shailasree, S., Kini, R. K., Muck, A., Mithöfer, A., and Shetty, S. H. (2010). Hydroxyproline-rich glycoproteins and plant defense. *J. Phytopathol.* 158, 585–593. doi: 10.1111/j.1439-0434.2010.01669.x
- Deng, Y. Y., Li, J. Q., Wu, S. F., Zhu, Y. P., Chen, Y. W., and He, F. C. (2006). Integrated NR database in protein annotation system and its localization. *Comput. Eng.* 32, 71–74. doi: 10.1109/INFOCOM.2006.241
- Dinesh-Kumar, S. P., and Baker, B. J. (2000). Alternatively spliced *N* resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proc. Natl. Acad. Sci. U.S.A.* 97, 1908–1913. doi: 10.1073/pnas.020367497
- Dixon, R. A., Achnine, L., Kota, P., Liu, C. J., Reddy, M. S. S., and Wang, L. (2002). The phenylpropanoid pathway and plant defence—a genomics perspective. *Mol. Plant Pathol.* 3, 371–390. doi: 10.1046/j.1364-3703.2002.00131.x
- Edens, R. M., Anand, S. C., and Bolla, R. I. (1995). Enzymes of the phenylpropanoid pathway in soybean infected with *Meloidogyne incognita* or *Heterodera glycines*. *J. Nematol.* 27, 292–303. doi: 10.1006/jipa.1995.1090
- Eitas, T. K., and Dangl, J. L. (2010). NB-LRR proteins: pairs, pieces, perception, partners, and pathways. *Curr. Opin. Plant Biol.* 13, 472–477. doi: 10.1016/j.pbi.2010.04.007
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Finotello, F., and Di Camillo, B. (2015). Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. *Brief. Funct. Genomics* 14, 130–142. doi: 10.1093/bfpg/elu035
- Foissac, S., and Sammeth, M. (2007). ASTALAVISTA: dynamic and flexible analysis of alternative splicing events in custom gene datasets. *Nucleic Acids Res.* 35, W297–W299. doi: 10.1093/nar/gkm311
- Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-sequencing the trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084
- Hirao, T., Fukatsu, E., and Watanabe, A. (2012). Characterization of resistance to pine wood nematode infection in *Pinus thunbergii* using suppression subtractive hybridization. *BMC Plant Biol.* 12:13. doi: 10.1186/1471-2229-12-13
- Hofmann, J., and Grundler, F. M. W. (2008). Starch as a sugar reservoir for nematode-induced syncytia. *Plant Signal. Behav.* 3, 961–962. doi: 10.4161/psb.6075
- Hofmann, J., Hess, P. H., Szakassits, D., Blöchl, A., Wiczorek, K., Daxböck-Horvath, S., et al. (2009). Diversity and activity of sugar transporters in nematode-induced root syncytia. *J. Exp. Bot.* 60, 3085–3095. doi: 10.1093/jxb/erp138
- Hofmann, J., Wiczorek, K., Blöchl, A., and Grundler, F. M. W. (2007). Sucrose supply to nematode-induced syncytia depends on the apoplasmic and symplasmic pathways. *J. Exp. Bot.* 58, 1591–1601. doi: 10.1093/jxb/erl285
- Hua, C., Li, C. J., Hu, Y. F., Mao, Y. Z., You, J., Wang, M. Z., et al. (2018). Identification of HG types of soybean cyst nematode *Heterodera glycines* and resistance screening on soybean genotypes in Northeast China. *J. Nematol.* 50, 41–50. doi: 10.21307/jofnem-2018-007
- Huang, M., Qin, R., Li, C., Liu, C., Jiang, Y., Yu, J., et al. (2021). Transgressive resistance to *Heterodera glycines* in chromosome segment substitution lines derived from susceptible soybean parents. *Plant Genome* 14:e20091. doi: 10.1002/tpg2.20091
- Huang, M., Qin, R., Li, C., Wang, M., Jiang, Y., Yu, J., et al. (2022). Response of soybean genotypes from Northeast China to *Heterodera glycines* races 4 and 5, and characterisation of *rhg1* and *Rhg4* genes for soybean resistance. *Nematology* 24, 333–345. doi: 10.1163/15685411-bja10134
- Ithal, N., Recknor, J., Nettleton, D., Hearne, L., Maier, T., Baum, T. J., et al. (2007). Parallel genome-wide expression profiling of host and pathogen during soybean cyst nematode infection of soybean. *Mol. Plant Microbe Interact.* 20, 293–305. doi: 10.1094/MPMI-20-3-0293
- Jiang, H. P., Bu, F. S., Tian, L. Z., Sun, Q. X., Bao, D. F., Zhao, X., et al. (2020). RNA-seq-based identification of potential resistance mechanism against the soybean cyst nematode (*Heterodera glycines*) HG Type 0 in soybean (*Glycine max*) cv. Dongnong L-204. *Crop Pasture Sci.* 71, 539–551. doi: 10.1071/CP20060
- John, S., Olas, J. J., and Mueller-Roeber, B. (2021). Regulation of alternative splicing in response to temperature variation in plants. *J. Exp. Bot.* 72, 6150–6163. doi: 10.1093/jxb/erab232
- Kandath, P. K., Itthal, N., Recknor, J., Maier, T., Nettleton, D., Baum, T. J., et al. (2011). The Soybean *Rhg1* locus for resistance to the soybean cyst nematode *Heterodera glycines* regulates the expression of a large number of stress- and defense-related genes in degenerating feeding cells. *Plant Physiol.* 155, 1960–1975. doi: 10.1104/pp.110.167536
- Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi: 10.1093/nar/gkh063
- Kang, W., Zhu, X., Wang, Y., Chen, L., and Duan, Y. (2018). Transcriptomic and metabolomic analyses reveal that bacteria promote plant defense during infection of soybean cyst nematode in soybean. *BMC Plant Biol.* 18:86. doi: 10.1186/s12870-018-1302-9
- Kasai, T., Inoue, M., Koshiba, S., Yabuki, T., Aoki, M., Nunokawa, E., et al. (2004). Solution structure of a BolA-like protein from *Mus musculus*. *Protein Sci.* 13, 545–548. doi: 10.1110/ps.03401004
- Khoei, M. A., Karimi, M., Karamian, R., Amini, S., and Soorni, A. (2021). Identification of the complex interplay between nematode-related lncRNAs and their target genes in *Glycine max* L. *Front. Plant Sci.* 12:779597. doi: 10.3389/fpls.2021.779597
- Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. *EMBO J.* 20, 3617–3622. doi: 10.1093/emboj/20.14.3617
- Klink, V. P., and Matthews, B. F. (2009). Emerging approaches to broaden resistance of soybean to soybean cyst nematode as supported by gene expression studies. *Plant Physiol.* 151, 1017–1022. doi: 10.1104/pp.109.144006
- Klink, V. P., Overall, C. C., Alkharouf, N. W., MacDonald, M. H., and Matthews, B. F. (2007). A time-course comparative microarray analysis of an incompatible and compatible response by *Glycine max* (soybean) to *Heterodera glycines* (soybean cyst nematode) infection. *Planta* 226, 1423–1447. doi: 10.1007/s00425-007-0581-4
- Kong, L., Zhang, Y., Ye, Z., Liu, X., Zhao, S., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 36, W345–W349. doi: 10.1093/nar/gkm391
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5:R7. doi: 10.1186/gb-2004-5-2-r7
- Lai, Z., Li, Y., Wang, F., Cheng, Y., Fan, B., Yu, J. Q., et al. (2011). *Arabidopsis* sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *Plant Cell* 23, 3824–3841. doi: 10.1105/tpc.111.090571
- Lalitha, S. (2000). Primer premier 5. *Biotech. Softw. Internet Rep.* 1, 270–272. doi: 10.1089/152791600459894
- Li, C., Xiang, X., Huang, Y., Zhou, Y., An, D., Dong, J., et al. (2020). Long-read sequencing reveals genomic structural variations that underlie creation of quality protein maize. *Nat. Commun.* 11:17. doi: 10.1038/s41467-019-14023-2
- Li, X., Qin, R., Du, Q., Cai, L., Hu, D., Du, H., et al. (2020). Knockdown of *GmVQ58* encoding a VQ motif-containing protein enhances soybean resistance to the common cutworm (*Spodoptera litura* Fabricius). *J. Exp. Bot.* 71, 3198–3210. doi: 10.1093/jxb/eraa095
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

- Li, J. W., Ma, W., Zeng, P., Wang, Z. Y., Geng, B., Yang, J. C., et al. (2015). LncTar: a tool for predicting the RNA targets of long noncoding RNAs. *Brief. Bioinform.* 16:806. doi: 10.1093/bib/bbu048
- Li, S., Chen, Y., Zhu, X., Wang, Y., Jung, K. H., Chen, L., et al. (2018). The transcriptomic changes of Huipizhi Heidou (*Glycine max*), a nematode-resistant black soybean during *Heterodera glycines* race 3 infection. *J. Plant Physiol.* 220, 96–104. doi: 10.1016/j.jplph.2017.11.001
- Li, X., Wang, X., Zhang, S., Liu, D., Duan, Y., and Dong, W. (2012). Identification of soybean microRNAs involved in soybean cyst nematode infection by deep sequencing. *PLoS One* 7:e39650. doi: 10.1371/journal.pone.0039650
- Li, Y., Qi, X., Chang, R., and Qiu, L. (2011). "Evaluation and utilization of soybean germplasm for resistance to cyst nematode in China," in *Soybean-Molecular Aspects of Breeding*, ed. A. Sudaric (London: InTech Open), 373–396. doi: 10.5772/14379
- Li, X., Wang, X., Zhang, S., Liu, D., Duan, Y., and Dong, W. (2011). Comparative profiling of the transcriptional response to soybean cyst nematode infection of soybean roots by deep sequencing. *Chin. Sci. Bull.* 56:1904. doi: 10.1007/s11434-011-4510-3
- Liu, S., Kandath, P. K., Lakhssassi, N., Kang, J., Colantonio, V., and Heinz, R. (2017). The soybean *GmSNAP18* gene underlies two types of resistance to soybean cyst nematode. *Nat. Commun.* 8:14822. doi: 10.1038/ncomms14822
- Liu, S., Kandath, P. K., Warren, S. D., Yeckel, G., Heinz, R., and Alden, J. (2012). A soybean cyst nematode resistance gene points to a new mechanism of plant resistance to pathogens. *Nature* 492, 256–260. doi: 10.1038/nature11651
- Mandadi, K. K., and Scholthof, K. B. (2015). Genome-wide analysis of alternative splicing landscapes modulated during plant-virus interactions in *Brachypodium distachyon*. *Plant Cell* 27, 71–85. doi: 10.1105/tpc.114.133991
- Mao, X. Z., Cai, T., Olyarchuk, J. G., and Wei, L. P. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics* 21, 3787–3793. doi: 10.1093/bioinformatics/bti430
- Martín, G., Márquez, Y., Mantica, F., Duque, P., and Irimia, M. (2021). Alternative splicing landscapes in *Arabidopsis thaliana* across tissues and stress conditions highlight major functional differences with animals. *Genome Biol.* 22:35. doi: 10.1186/s13059-020-02258-y
- Matsukura, S., Mizoi, J., Yoshida, T., Todaka, D., Ito, Y., Maruyama, K., et al. (2010). Comprehensive analysis of rice *DREB2*-type genes that encode transcription factors involved in the expression of abiotic stress-responsive genes. *Mol. Genet. Genom.* 283, 185–196. doi: 10.1007/s00438-009-0506-y
- Matsye, P. D., Kumar, R., Hosseini, P., Jones, C. M., Tremblay, A., Alkharouf, N. W., et al. (2011). Mapping cell fate decisions that occur during soybean defense responses. *Plant Mol. Biol.* 77, 513–528. doi: 10.1007/s11103-011-9828-3
- Mazarei, M., Liu, W., Al-Ahmad, H., Arelli, P. R., Pantalone, V. R., and Stewart, C. N. Jr. (2011). Gene expression profiling of resistant and susceptible soybean lines infected with soybean cyst nematode. *Theor. Appl. Genet.* 123, 1193–1206. doi: 10.1007/s00122-011-1659-8
- Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, 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
- Mine, A., Seyfferth, C., Kracher, B., Berens, M. L., Becker, D., and Tsuda, K. (2018). The defense phytohormone signaling network enables rapid, high-amplitude transcriptional reprogramming during effector-triggered immunity. *Plant Cell* 30, 1199–1219. doi: 10.1105/tpc.17.00970
- Miraeiz, E., Chaiprom, U., Afsharifard, A., Karegar, A., Drnevich, J. M., and Hudson, M. E. (2020). Early transcriptional responses to soybean cyst nematode HG Type 0 show genetic differences among resistant and susceptible soybeans. *Theor. Appl. Genet.* 133, 87–102. doi: 10.1007/s00122-019-03442-w
- Mitchum, M. G., and Baum, T. J. (2008). "Genomics of the soybean cyst nematode-soybean interaction," in *Genetics and Genomics of Soybean*, ed. G. Stacey (New York, NY: Springer), 321–341. doi: 10.1007/978-0-387-72299-3_17
- Mitchum, M. G., Wrather, J. A., Heinz, R. D., Shannon, J. G., and Danekas, G. (2007). Variability in distribution and virulence phenotypes of *Heterodera glycines* in Missouri during 2005. *Plant Dis.* 91, 1473–1476. doi: 10.1094/PDIS-91-11-1473
- Monaghan, J., and Zipfel, C. (2012). Plant pattern recognition receptor complexes at the plasma membrane. *Curr. Opin. Plant Biol.* 15, 349–357. doi: 10.1016/j.pbi.2012.05.006
- Mosser, A., Aller, I., Wagner, S., Nietzel, T., Przybyla-Toscano, J., Mühlenhoff, U., et al. (2015). The mitochondrial monothiol glutaredoxin S15 is essential for iron-sulfur protein maturation in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13735–13740. doi: 10.1073/pnas.1510835112
- Naveed, Z. A., Wei, X., Chen, J., Mubeen, H., and Ali, G. S. (2020). The PTI to ETI continuum in phytophthora-plant interactions. *Front. Plant Sci.* 11:593905. doi: 10.3389/fpls.2020.593905
- Neupane, S., Mathew, F. M., Varenhorst, A. J., and Nepal, M. P. (2019). Transcriptome profiling of interaction effects of soybean cyst nematodes and soybean aphids on soybean. *Sci. Data* 6:133. doi: 10.1038/s41597-019-0140-4
- Niblack, T. L., Colgrove, A. L., Colgrove, K., and Bond, J. P. (2008). Shift in virulence of soybean cyst nematode is associated with use of resistance from PI 88788. *Plant Health Prog.* 9:29. doi: 10.1094/PHP-2008-0118-01-RS
- Niblack, T. L., Lambert, K. N., and Tylka, G. L. (2006). A model plant pathogen from the kingdom animalia: *Heterodera glycines*, the soybean cyst nematode. *Annu. Rev. Phytopathol.* 44, 283–303. doi: 10.1146/annurev.phyto.43.040204.140218
- Pandey, V. P., Awasthi, M., Singh, S., Tiwari, S., and Dwivedi, U. N. (2017). A comprehensive review on function and application of plant peroxidases. *Anal. Biochem.* 6:308. doi: 10.4172/2161-1009.100030
- Perruc, E., Charpentier, M., Ramirez, B. C., Jauneau, A., Galaud, J. P., Ranjeva, R., et al. (2004). A novel calmodulin-binding protein functions as a negative regulator of osmotic stress tolerance in *Arabidopsis thaliana* seedlings. *Plant J.* 38, 410–420. doi: 10.1111/j.1365-3113X.2004.02062.x
- Puthoff, D. P., Ehrenfried, M. L., Vinyard, B. T., and Tucker, M. L. (2007). GeneChip profiling of transcriptional responses to soybean cyst nematode, *Heterodera glycines*, colonization of soybean roots. *J. Exp. Bot.* 58, 3407–3418. doi: 10.1093/jxb/erm211
- Qin, L., Wang, M., Zuo, J., Feng, X., Liang, X., Wu, Z., et al. (2015). Cytosolic BoLA plays a repressive role in the tolerance against excess iron and MV-induced oxidative stress in plants. *PLoS One* 10:e0124887. doi: 10.1371/journal.pone.0124887
- Qiu, J. L., Fiil, B. K., Petersen, K., Nielsen, H. B., Botanga, C. J., Thorgrimsen, S., et al. (2008). *Arabidopsis* MAP kinase 4 regulates gene expression through transcription factor release in the nucleus. *EMBO J.* 27, 2214–2221. doi: 10.1038/emboj.2008.147
- Sadek, J., Omer, A., Hall, D., Ashour, K., and Gallouzi, I. E. (2019). Alternative polyadenylation and the stress response. *Wiley Interdiscip. Rev. RNA* 10:e1540. doi: 10.1002/wrna.1540
- Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P., and De Bolle, M. F. (2008). Plant pathogenesis-related (PR) proteins: a focus on PR peptides. *Plant Physiol. Biochem.* 46, 941–950. doi: 10.1016/j.plaphy.2008.06.011
- Severing, E. I., Van Dijk, A. D., Morabito, G., Busscher-Lange, J., Immink, R. G., and Van Ham, R. C. (2012). Predicting the impact of alternative splicing on plant MADS domain protein function. *PLoS One* 7:e30524. doi: 10.1371/journal.pone.0030524
- 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
- Singh, R. R., Chinnasri, B., De Smet, L., Haeck, A., Demeestere, K., Van Cutsem, P., et al. (2019). Systemic defense activation by COS-OGA in rice against root-knot nematodes depends on stimulation of the phenylpropanoid pathway. *Plant Physiol. Biochem.* 142, 202–210. doi: 10.1016/j.plaphy.2019.07.003
- Song, H., Wang, P., Hou, L., Zhao, S., Zhao, C., Xia, H., et al. (2016). Global analysis of WRKY genes and their response to dehydration and salt stress in soybean. *Front. Plant Sci.* 7:9. doi: 10.3389/fpls.2016.00009
- Song, W., Qi, N., Liang, C., Duan, F., and Zhao, H. (2019). Soybean root transcriptome profiling reveals a nonhost resistant response during *Heterodera glycines* infection. *PLoS One* 14:e0217130. doi: 10.1371/journal.pone.0217130
- Stock, A. M., Robinson, V. L., and Goudreau, P. N. (2000). Two-component signal transduction. *Annu. Rev. Biochem.* 69, 183–215. doi: 10.1146/annurev.biochem.69.1.183

- Sun, L., Luo, H., Bu, D., Zhao, G., Yu, K., Zhang, C., et al. (2013). Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Res.* 41:e166. doi: 10.1093/nar/gkt646
- Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. (2012). Alternative splicing in plants-coming of age. *Trends Plant Sci.* 17, 616–623. doi: 10.1016/j.tplants.2012.06.001
- Szczeniński, M. W., Rosikiewicz, W., and Makalowska, I. (2015). CANTATAdB: a collection of plant long non-coding RNAs. *Plant Cell Physiol.* 57:e8. doi: 10.1093/pcp/pcv201
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Thirupathi, D. (2020). SPLICE in the seeds: integration of abscisic acid and light signaling in *Arabidopsis*. *Plant Physiol.* 183, 445–446. doi: 10.1104/pp.20.00361
- Trabucco, G. M., Matos, D. A., Lee, S. J., Saathoff, A. J., and Hazen, S. P. (2013). Functional characterization of cinnamyl alcohol dehydrogenase and caffeic acid O-methyltransferase in *Brachypodium distachyon*. *BMC Biotechnol.* 13:61. doi: 10.1186/1472-6750-13-61
- Tu, Z., Shen, Y., Wen, S., Liu, H., Wei, L., and Li, H. (2021). A Tissue-Specific landscape of alternative polyadenylation, lncRNAs, TFs, and gene co-expression networks in *Liriodendron chinense*. *Front. Plant Sci.* 12:705321. doi: 10.3389/fpls.2021.705321
- Urquiaga, M. C. O., Thiebaut, F., Hemery, A. S., and Ferreira, P. C. G. (2021). From trash to luxury: the potential role of plant lncRNA in DNA methylation during abiotic stress. *Front. Plant Sci.* 11:603246. doi: 10.3389/fpls.2020.603246
- Wan, J., Vuong, T., Jiao, Y., Joshi, T., Zhang, H., Xu, D., et al. (2015). Whole-genome gene expression profiling revealed genes and pathways potentially involved in regulating interactions of soybean with cyst nematode (*Heterodera glycines* Ichinohe). *BMC Genomics* 16:148. doi: 10.1186/s12864-015-1316-8
- Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J. P., and Li, W. (2013). CPAT: coding-potential assessment tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41:e74. doi: 10.1093/nar/gkt006
- Wang, M., Wang, P., Liang, F., Ye, Z., Li, J., Shen, C., et al. (2018). A global survey of alternative splicing in allopolyploid cotton: landscape, complexity and regulation. *New Phytol.* 217, 163–178. doi: 10.1111/nph.14762
- Wang, X., Zhang, H., Sun, G., Jin, Y., and Qiu, L. (2014). Identification of active VQ motif-containing genes and the expression patterns under low nitrogen treatment in soybean. *Gene* 543, 237–243. doi: 10.1016/j.gene.2014.04.012
- Wang, Y., Jiang, Z., Li, Z., Zhao, Y., Tan, W., Liu, Z., et al. (2019). Genome-wide identification and expression analysis of the VQ gene family in soybean (*Glycine max*). *PeerJ* 7:e7509. doi: 10.7717/peerj.7509
- Xie, S., Leung, A. W., Zheng, Z., Zhang, D., Xiao, C., Luo, R., et al. (2021). Applications and potentials of nanopore sequencing in the (epi)genome and (epi)transcriptome era. *Innovation* 2:100153. doi: 10.1016/j.xinn.2021.100153
- Xu, X., Zhang, L., Zhao, W., Fu, L., Han, Y., Wang, K., et al. (2021). Genome-wide analysis of the serine carboxypeptidase-like protein family in *Triticum aestivum* reveals *TaSCPL184-6D* is involved in abiotic stress response. *BMC Genomics* 22:350. doi: 10.1186/s12864-021-07647-6
- Yan, C., Wang, Y., Lyu, T., Hu, Z., Ye, N., Liu, W., et al. (2021). Alternative polyadenylation in response to temperature stress contributes to gene regulation in *Populus trichocarpa*. *BMC Genomics* 22:53. doi: 10.1186/s12864-020-07353-9
- Yang, J., Lv, W., Shao, L., Fu, Y., Liu, H., Yang, C., et al. (2021). PacBio and Illumina RNA sequencing identify alternative splicing events in response to cold stress in two poplar species. *Front. Plant Sci.* 12:737004. doi: 10.3389/fpls.2021.737004
- Yang, Y., Zhou, Y., Chi, Y., Fan, B., and Chen, Z. (2017). Characterization of soybean WRKY gene family and identification of soybean WRKY genes that promote resistance to soybean cyst nematode. *Sci. Rep.* 7:17804. doi: 10.1038/s41598-017-18235-8
- Yao, T., Feng, K., Xie, M., Barros, J., Tschaplinski, T. J., Tuskan, G. A., et al. (2021). Phylogenetic occurrence of the phenylpropanoid pathway and lignin biosynthesis in plants. *Front. Plant Sci.* 12:704697. doi: 10.3389/fpls.2021.704697
- Ye, C., Zhou, Q., Wu, X., Ji, G., and Li, Q. Q. (2019). Genome-wide alternative polyadenylation dynamics in response to biotic and abiotic stresses in rice. *Ecotoxicol. Environ. Saf.* 183:109485. doi: 10.1016/j.ecoenv.2019.109485
- Yeh, H. S., and Yong, J. (2016). Alternative polyadenylation of mRNAs: 3'-untranslated region matters in gene expression. *Mol. Cells* 39, 281–285. doi: 10.14348/molcells.2016.0035
- Young, M. D., Wakefield, M. J., and Smyth, G. K. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biol.* 11:R14. doi: 10.1186/gb-2010-11-2-r14
- Yu, Y., Wang, N., Hu, R., and Xiang, F. (2016). Genome-wide identification of soybean WRKY transcription factors in response to salt stress. *Springerplus* 5:920. doi: 10.1186/s40064-016-2647-x
- Yu, Y., Zhou, Y. F., Feng, Y. Z., He, H., Lian, J. P., Yang, Y. W., et al. (2020). Transcriptional landscape of pathogen-responsive lncRNAs in rice unveils the role of ALEX1 in jasmonate pathway and disease resistance. *Plant Biotechnol. J.* 18, 679–690. doi: 10.1111/pbi.13234
- Yuan, M., Ngou, B., Ding, P., and Xin, X. F. (2021). PTI-ETI crosstalk: an integrative view of plant immunity. *Curr. Opin. Plant Biol.* 62:102030. doi: 10.1016/j.pbi.2021.102030
- Zeng, H., Xu, L., Singh, A., Wang, H., Du, L., and Poovaiah, B. W. (2015). Involvement of calmodulin and calmodulin-like proteins in plant responses to abiotic stresses. *Front. Plant Sci.* 6:600. doi: 10.3389/fpls.2015.00600
- Zhang, G., Sun, M., Wang, J., Lei, M., Li, C., and Zhao, D. (2019). PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. *Plant J.* 97, 296–305. doi: 10.1111/tpl.14120
- Zhang, H. Y., Kjemtrup-Lovelace, S., Li, C. B., Luo, Y., Chen, L. P., and Song, B. H. (2017). Comparative RNA-seq analysis uncovers a complex regulatory network for soybean cyst nematode resistance in wild soybean (*Glycine soja*). *Sci. Rep.* 7:9699. doi: 10.1038/s41598-017-09945-0
- Zhang, J., Gu, H., Dai, H., Zhang, Z., and Miao, M. (2020). Alternative polyadenylation of the stacyose synthase gene mediates source-sink regulation in cucumber. *J. Plant Physiol.* 245:153111. doi: 10.1016/j.jplph.2019.153111
- Zhao, L., Zhang, H., Kohnen, M. V., Prasad, K., Gu, L., and Reddy, A. (2019). Analysis of transcriptome and epitranscriptome in plants using PacBio iso-seq and Nanopore-based direct RNA sequencing. *Front. Genet.* 10:253. doi: 10.3389/fgene.2019.00253
- Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014
- Zhou, X., Lindsay, H., and Robinson, M. D. (2014). Robustly detecting differential expression in RNA sequencing data using observation weights. *Nucleic Acids Res.* 42:e91. doi: 10.1093/nar/gku310

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 Huang, Jiang, Qin, Jiang, Chang, Tian, 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.



Soybean F-Box-Like Protein GmFBL144 Interacts With Small Heat Shock Protein and Negatively Regulates Plant Drought Stress Tolerance

Keheng Xu¹, Yu Zhao¹, Yan Zhao¹, Chen Feng¹, Yinhe Zhang¹, Fawei Wang¹, Xiaowei Li¹, Hongtao Gao^{2,3}, Weican Liu¹, Yan Jing^{2,3}, Rachit K. Saxena⁴, Xianzhong Feng⁵, Yonggang Zhou^{2,3*} and Haiyan Li^{1,2,3*}

¹ College of Life Sciences, Jilin Agricultural University, Changchun, China, ² College of Tropical Crops, Sanya Nanfan Research Institute, Hainan University, Haikou, China, ³ Hainan Yazhou Bay Seed Laboratory, Sanya, China, ⁴ International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Hyderabad, India, ⁵ Key Laboratory of Soybean Molecular Design Breeding, Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, China

OPEN ACCESS

Edited by:

Umezuruike Linus Opara,
Stellenbosch University, South Africa

Reviewed by:

Igor Obuchowski,
University of Gdańsk, Poland
Zhao-Shi Xu,
Institute of Crop Sciences (CAAS),
China

*Correspondence:

Yonggang Zhou
yonggang0408@163.com
Haiyan Li
hly199@163.com

Specialty section:

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

Received: 27 November 2021

Accepted: 28 April 2022

Published: 02 June 2022

Citation:

Xu K, Zhao Y, Zhao Y, Feng C,
Zhang Y, Wang F, Li X, Gao H, Liu W,
Jing Y, Saxena RK, Feng X, Zhou Y
and Li H (2022) Soybean F-Box-Like
Protein GmFBL144 Interacts With
Small Heat Shock Protein
and Negatively Regulates Plant
Drought Stress Tolerance.
Front. Plant Sci. 13:823529.
doi: 10.3389/fpls.2022.823529

The *F-box* gene family is one of the largest gene families in plants. These genes regulate plant growth and development, as well as biotic and abiotic stress responses, and they have been extensively researched. Drought stress is one of the major factors limiting the yield and quality of soybean. In this study, bioinformatics analysis of the soybean *F-box* gene family was performed, and the role of soybean *F-box-like* gene *GmFBL144* in drought stress adaptation was characterized. We identified 507 *F-box* genes in the soybean genome database, which were classified into 11 subfamilies. The expression profiles showed that *GmFBL144* was highly expressed in plant roots. Overexpression of *GmFBL144* increased the sensitivity of transgenic *Arabidopsis* to drought stress. Under drought stress, the hydrogen peroxide (H₂O₂) and malonaldehyde (MDA) contents of transgenic *Arabidopsis* were higher than those of the wild type (WT) and empty vector control, and the chlorophyll content was lower than that of the control. Y2H and bimolecular fluorescence complementation (BiFC) assays showed that *GmFBL144* can interact with GmSHSP. Furthermore, our results showed that *GmFBL144* can form SCF^{FBL144} (E3 ubiquitin ligase) with GmSkp1 and GmCullin1. Altogether, these results indicate that the soybean *F-box-like* protein *GmFBL144* may negatively regulate plant drought stress tolerance by interacting with SHSP. These findings provide a basis for molecular genetics and breeding of soybean.

Keywords: soybean, F-box protein, drought stress, Skp1-Cullin1-F-box (SCF) complex, segmental duplication, protein interaction

INTRODUCTION

Abiotic stress is the primary factor limiting plant growth and crop yield. Abiotic stresses include drought, saline-alkali, high/low temperature, and metal stress, of which drought stress is the most common stress (Baldoni et al., 2015; Gong et al., 2020). Plants have evolved multiple strategies to deal with drought stress. Common strategies include reducing water loss, maintaining chlorophyll content, and reducing reactive oxygen (Ritchie et al., 1990; Zafari et al., 2020).

Molecular breeding has long been expected to improve crop drought tolerance. Thus far, many genes involved in drought stress regulation have been identified. For instance, *ZmVPP1*, a vacuolar-type H⁺ pyrophosphatase gene, can improve the drought tolerance of transgenic maize by enhancing root development and photosynthetic efficiency (Wang X. et al., 2016). Tubby-like F-box protein 8 (SITLTP8) enhances plant drought tolerance by reducing water loss to improve water-use efficiency (Li et al., 2020b). Additionally to the above positive regulatory genes, some negative regulatory genes have been found. For example, MdSE reduces the expression of *MdNCED3* by negatively regulating the MdMYB88 and MdMYB124 transcription factors (MdMYB88, MdMYB124, *MdNCED3*; positive regulators of drought resistance) to reduce the drought tolerance of apples (Li et al., 2020c). Although many genes responding drought stress have been investigated, the molecular network of plant responses to drought stress is still imperfect.

The ubiquitin-proteasome system (UPS) and molecular chaperone system play important roles in plant responses to drought stress. Protein degradation mediated by UPS is an important post-translation regulation mechanism, which includes ubiquitin, ubiquitin-activating enzymes (E1s), ubiquitin conjugating enzymes (E2s), ubiquitin ligase enzymes (E3s), and 26S proteasomes (Sadanandom et al., 2012; Xia et al., 2020; Ban and Estelle, 2021). Ubiquitin is activated by E1 in the presence of ATP, and is then transferred to the cysteine residue of E2. Then, E3 transfers ubiquitin to the lysine residue of the substrate, and finally 26S proteasome degrades the ubiquitinated substrate. In this pathway, E3s are responsible for substrate recognition and substrate ubiquitination. Research has found that there are thousands of E3s in plants, which can be classified as many different types, of which RING E3s are the most abundant (Zheng and Shabek, 2013; Morreale and Walden, 2016). SCF is a well characterized RING E3 ubiquitin ligase, which contains four subunits: Cullin1 (CUL1), Ring-box protein (Rbx), S-Phase kinase associated protein 1 (Skp1), and F-box protein. Cullin1, Rbx, and Skp1 interact to form a core scaffold with ligase activity, and the F-box proteins are responsible for substrate recognition (Zheng et al., 2016; Zhang S. et al., 2019).

Recently, large numbers of F-box proteins were identified. Especially in plants, there are hundreds of F-box proteins. Their common feature is that the N-terminal contains a relatively conservative F-box domain that interacts with Skp1 and CUL1 to form SCF. Besides the F-box domain, there are other variable domains in the C-terminal. The C-terminal domains are responsible for substrate recognition and also provide a foundation for subfamily classification of F-box proteins. The diversity of F-box proteins can help SCF distinguish and recruit multiple substrates. Therefore, it is not surprising that F-box proteins can regulate many physiological processes. For example, Carbonnel et al. (2020) found that F-box protein MAX2 can inhibit the growth of primary roots and promote the growth of root hairs by increasing the content of ethylene through the karrikins signaling pathway. Another case in point is that ORE9, an F-box protein, can regulate leaf senescence through the ubiquitin-proteasome pathway (Woo et al., 2001;

Zhang et al., 2016). In addition, in recent years, a growing number of F-box proteins involved in abiotic stress responses have been studied, e.g., drought, salt, ion, and low temperature stresses (Lim et al., 2019; Zhang Y. et al., 2019; Jie et al., 2020; Venkatesh et al., 2020).

Presently, most research on F-box protein function are from other plants, and research on soybean F-box proteins is relatively limited. Soybean is an important food crop and oil crop, and its yield is seriously affected by drought stress. Therefore, it is necessary to study the function of GmF-box proteins in drought. In this study, we performed functional characterization of *GmFBL144* in drought stress adaptation. The results showed that overexpression of *GmFBL144* significantly reduced plant drought tolerance. Under drought stress, the overexpression lines had higher hydrogen peroxide (H₂O₂) and malonaldehyde (MDA) content, lower chlorophyll content, and a higher water loss rate compared to the controls [wild type (WT) and empty vector]. Furthermore, our results showed that GmFBL144 can form SCF^{FBL144} (E3 ubiquitin ligase) with GmSkp1 and GmCulin1, and GmFBL144 can interact with GmHSP.

MATERIALS AND METHODS

Identification and Bioinformatics Analysis of *GmF-Box* Genes

The whole genome protein sequence of soybean (Glycine max Wm82.a2.v1) was downloaded from the plant genome database¹. Simultaneously, the hidden Markov model of F-box domain (PF00646) and F-box-like domain (PF12937) were downloaded from the Pfam database². The proteins sequences were searched by conserved domain using the BLAST method in the protein database. The genes of the two families were merged and then identified using the NCBI database (*E*-value cutoff 1.0). The genes without the F-box domain and F-box-like domain were deleted.

Phylogenetic analysis was inferred using the Maximum Likelihood method based on the Poisson correction model by MEGA7 (1000 bootstrap replicates for detection reliability). The annotation file GFF3 for soybean was downloaded from the Phytozome database, and visualized using TBtools. Gene duplication analyses were performed with the MCScanX (Wang et al., 2013).

A 2,000 bp DNA sequence upstream of *GmFBL144* gene was download from the soybean genome database, and the promoter elements were analyzed using the PlantCARE website³ (Lescot et al., 2002).

Plant Materials and Growth Conditions

The soybean genotype Williams 82 was cultivated in the field. Roots, stems, leaves, flowers, and embryos of different stages were collected and stored at −80°C for RNA extraction for quantitative RT-PCR assay. *Arabidopsis thaliana* were cultivated

¹<http://phytozome.jgi.doe.gov>

²<http://pfam.xfam.org>

³<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>

in a controlled-climate room with 16/8 h light/dark cycle at 23°C and 60% relative humidity.

Generation of Transgenic *Arabidopsis*

The full-length coding sequence (CDS) of *GmGBL144* was inserted after the CaMV35S promoter, resulting in overexpression of recombinant vectors. After sequencing, the fusion constructs were transformed into *Agrobacterium tumefaciens* (GV3101) and then transformed into *Arabidopsis* (Col-0) plants using the floral dip (Clough and Bent, 2010). T1 seedlings were screened with 1% glufosinate and then re-confirmed using genomic PCR. After screening of the separation ratio and high expression, three homozygous T3 transgenic lines (L16, L17, and L19) were selected for phenotypic experiments.

Stress Treatments

For rapid assay of the function of *GmFBL144* in the drought response, the full-length CDS of *GmGBL144* was inserted in pYES2 vector. The fusion construct was transformed into INVSc1 yeast competent cells using the PEG/LiAc-mediated method and was selected for uracil prototrophy. The function assay was implemented on YPD plate with 1.0 M mannitol. The cultures were grown for 3–7 days at 30°C.

After harvesting transgenic *Arabidopsis* seeds, we planted the seeds on 1/2 MS medium. These seedlings were transferred to plate of 1/2 MS with 250 mM mannitol. The root length and fresh weight were measured 2 weeks later. To further confirm the function of *GmFBL144*, the transgenic *Arabidopsis* were subjected to drought (no irrigation for up to 10 and 20 days) during the rosette stage. Rehydration was implemented after 20 days of drought. *Nicotiana benthamiana* were cultivated in a controlled-climate room with 14/10 h light/dark cycle at 26°C and 50% relative humidity.

Quantitative Real-Time PCR Analysis

Total RNA was extracted from the different soybean tissues using RNAiso Plus (Takara, DaLian, China). The cDNA was synthesized using the PrimeScriptTM RT reagent kit (Takara). Quantitative real-time PCR (qRT-PCR) was performed using TB Green Premix Ex TaqTM II (Takara). The relative expression was calculated using $2^{-\Delta C_t}$ (Li et al., 2021). The primers of qRT-PCR are listed in **Supplementary Table 1**.

Subcellular Localization of GmFBL144 Proteins

The full-length *GmFBL144* gene without stop codon was inserted after the CaMV35S promoter to generate a 35S:GmFBL144-GFP fusion construct. The fusion construct was transformed into *Agrobacterium tumefaciens* (GV3101, pSoup-p19) and then transferred into epidermal *Nicotiana benthamiana* cells for transient assays. The fluorescent signals of green fluorescent protein (GFP) were observed using a laser confocal microscope (TCS SP8, Leica, Germany).

Protein Interaction Assay

The prey cDNA library was constructed using CloneMiner II cDNA Library Construction Kit (Invitrogen: A11180) for

screening interacting proteins. The full-length *GmFBL144* gene was inserted into pGBKT7 to generate a bait vector. The positive clones were selected on SD/-His/-Leu/-Trp/-Ade/X- α -gal plates by yeast two-hybrid (Y2H) (Chen et al., 2013), and then these clones were identified by sequencing and Blastx. All protocols were carried out in strict accordance with the manufacturer's instructions.

The protein interactions were further verified using a bimolecular fluorescence complementation (BiFC) assay. GmFBL144 and GmsHSP were fused to nYFP and cYFP, respectively. The fusion constructs were transformed into *Agrobacterium tumefaciens* (GV3101, pSoup-p19) and then co-transferred into epidermal *Nicotiana benthamiana* cell for transient assays (Lu et al., 2009). The fluorescent signals of yellow fluorescent protein (YFP) were observed using a laser confocal microscope (TCS SP8, Leica, Germany).

The full-length *GmCullin1* gene (Glyma.17G025200) was inserted into pGADT7 and pGBKT7. The full-length *GmSkp1* gene was inserted into pGADT7. The full-length *GmFBL144* and *GmSkp1* genes (Glyma.11G079600) were inserted into two distinct multiple cloning sites of the pBridge vector simultaneously. The pairs were co-transformed into Y2H gold yeast competent cells using the PEG/LiAc-mediated method. The interaction between bait and prey proteins were identified on SD/-leu/-Trp and SD/-Leu/-Trp/-His/-Ade/X- α -gal media.

Determination of Chlorophyll, Malonaldehyde, and Hydrogen Peroxide Content

Chlorophyll content was assessed following previously used methods (Jie et al., 2020). The contents of MDA and H₂O₂ were determined through the corresponding test kits (Nanjing Jiancheng Bioengineering Institute, Nanjing, China).

Statistical Analysis

The results were analyzed using GraphPad Prism 5.0 software. The data are expressed as means \pm SD. All experiments were repeated thrice. The statistical significance was considered at the $P < 0.05$ or $P < 0.01$ level, as revealed by *t*-tests.

RESULTS

Identification and Bioinformatics Analysis of GmF-Box Genes

The *F-box* genes belong to a supergene family, which is a cluster of genes produced by duplication and mutation from an ancestor. Gm19G196400 was defined as an *F-box-like* gene belonging to the *F-box* gene family. To understand Gm19G196400, we performed bioinformatics analysis of the soybean *F-box* gene family. We identified 507 *F-box* genes in the soybean genome database, and we renamed these genes according to the N-terminal domain and chromosomal distribution. Our objective gene, Gm19G196400, is named *GmFBL144* (**Supplementary Table 2**).

F-box proteins contain other domains at C-terminal besides N-terminal relatively conserved F-box domain. In order to

research conveniently, the whole F-box protein family was classified into 11 subfamilies (**Supplementary Table 2**) and the GmFBL144 belongs to the FBO subfamily (F-box proteins with C-terminal other known domain). The phylogenetic tree analysis showed that most members of the same subfamily tended to cluster in the same evolutionary branch (**Supplementary Table 2** and **Supplementary Figure 1**).

The distribution of *GmF-box* genes on chromosomes was visualized (**Figure 1A**). The number of *F-box* genes on chromosome 8 was the largest (48 *F-box* genes), and on chromosome 1 was the lowest (10 *F-box* genes). There was no significant correlation between *F-box* gene number on a chromosome and chromosome length. Gene duplication is the source of evolutionary innovation and main factor in gene family expansion. In this study, gene duplication analysis was performed. 307 WGD/segmental duplication *GmF-box* genes, corresponding to 192 duplicated gene pairs in the entire *GmF-box* gene family, were identified in soybean; 109 dispersed duplication *GmF-box* genes, 44 proximal duplication *GmF-box* genes, and 26 tandem duplication *GmF-box* genes corresponded to 37, 17, and 9 duplicated genes pairs in the *GmF-box* gene family, respectively (**Figure 1B**). The results of the gene duplication analysis revealed that the *GmF-box* gene family expansion was largely a result of WGD/segmental duplication.

In addition, we analyzed the gene structure and promoter regions of *GmFBL144* and *GmFBL25* that showed the closest genetic relationship with *GmFBL144* (**Figures 2A,B** and **Supplementary Figure 1**). The results showed that *GmFBL144* had a similar structure to *GmFBL25* and adds a NleF-casp-inhib domain (**Figure 2B**). Furthermore, we found that there are four drought response elements (MYB motif, C/TAACNA/G) and a stress response element (TC-rich repeats, ATTCTCTAAC) in the promoter region of *GmFBL144*, whereas *GmFBL25* only had an MYB motif (**Figure 2C**). These findings suggest that *GmFBL144* may be involved in drought regulation. We also found that the relative expression of *GmFBL144* was high in roots (**Figure 2D**). Roots are the main organ of external environment perception. Therefore, we conjecture that *GmFBL144* may be involved in the drought stress response.

Overexpression of *GmFBL144* Increased Drought Stress Sensitivity in Transgenic *Arabidopsis* Seedlings

There are multiple drought response elements in the promoter region of *GmFBL144* and our transcriptome data of soybean (Zhou et al., 2020) showed that the *GmFBL144* was down-regulated under drought stress (**Supplementary Figure 2**). Accordingly, we studied the function of *GmFBL144* in drought stress. In this study, the yeast transient expression system was used. The results showed that the heterologous expression of *GmFBL144* yeast was undergrown compared with the vector control under drought stress conditions (**Figure 3A**). Next, we carried out the *Arabidopsis* plate drought stress study. Mannitol (250 mM) was used to simulate drought stress in 1/2 MS medium. The results showed that after 10 days of simulated drought stress, the primary root length of *GmFBL144*-overexpression

Arabidopsis was shorter and the fresh weight was lighter than that of controls (WT and empty vector) (**Figures 3B–D**).

Similarly, the transgenic *Arabidopsis* showed drought sensitivity under soil drought stress. Under normal conditions, there was no significant difference between controls (WT and empty vector) and the *GmFBL144*-overexpression transgenic lines. After 2 weeks without watering, the foliar loss of *GmFBL144*-overexpression transgenic lines was more serious than that in the controls (WT and empty vector) (**Figure 4A**). The chlorophyll content of *GmFBL144*-overexpression transgenic lines was lower, and the MDA and H₂O₂ content were higher than those of the controls (WT and empty vector) (**Figures 4B–G**). After 20 days without watering, the *GmFBL144*-overexpression transgenic lines were almost dead; however, the controls (WT and empty vector) were still alive. After 3 days of rehydration, the controls (WT and empty vector) recovered but the *GmFBL144*-overexpression transgenic lines could not (**Figure 4A**). These results indicate that *GmFBL144* increased plant drought sensitivity through damaging the system of scavenging activated oxygen.

Subcellular Localization of GmFBL144

Proteins are the most important biomacromolecules in organisms, and the major performers of life activities. Mature proteins can exert different biological functions in specific subcellular organelles. Thereby, the function of a protein is not only related to its structure but also to its subcellular localization. The subcellular localization prediction websites^{4, 5} forecast that the GmFBL144 protein could be located on the nucleus and chloroplast. At the same time, we investigated the subcellular localization of GmFBL144 using the transient transformation system of tobacco leaves. The fluorescence signal of the fusion protein was mainly located on the nucleus, and a small part also exist on the cell membrane (**Figure 5**).

Identification of the Formation Mechanism of the SCF^{FBL144} Complex

Studies have shown that most F-box proteins perform functions by forming SCF complexes (Kipreos and Pagano, 2000; Ao et al., 2020). To verify whether GmFBL144 can form SCF complex, we evaluated the interactions among Cullin1 (CUL1), Skp1, and GmFBL144 using Y2H assay. The results showed that GmFBL144 can interact with GmSkp1, cannot interact with GmCUL1; GmCUL1 can interact with GmSkp1 but is weak; the complex of GmFBL144-GmSkp1 can interact with GmCUL1. These results indicated that GmFBL144 is a key subunit of SCF^{FBL144}, and GmFBL144 can promote the combination of GmSkp1 and GmCUL1 (**Figure 6**).

Screening and Identification of GmFBL144 Interacting Proteins

In order to further research the mechanism, we screened the interacting proteins through the cDNA library. Because

⁴<https://www.genscript.com/wolf-psort.html>

⁵<http://www.csbio.sjtu.edu.cn/bioinf/Cell-PLoc-2/>

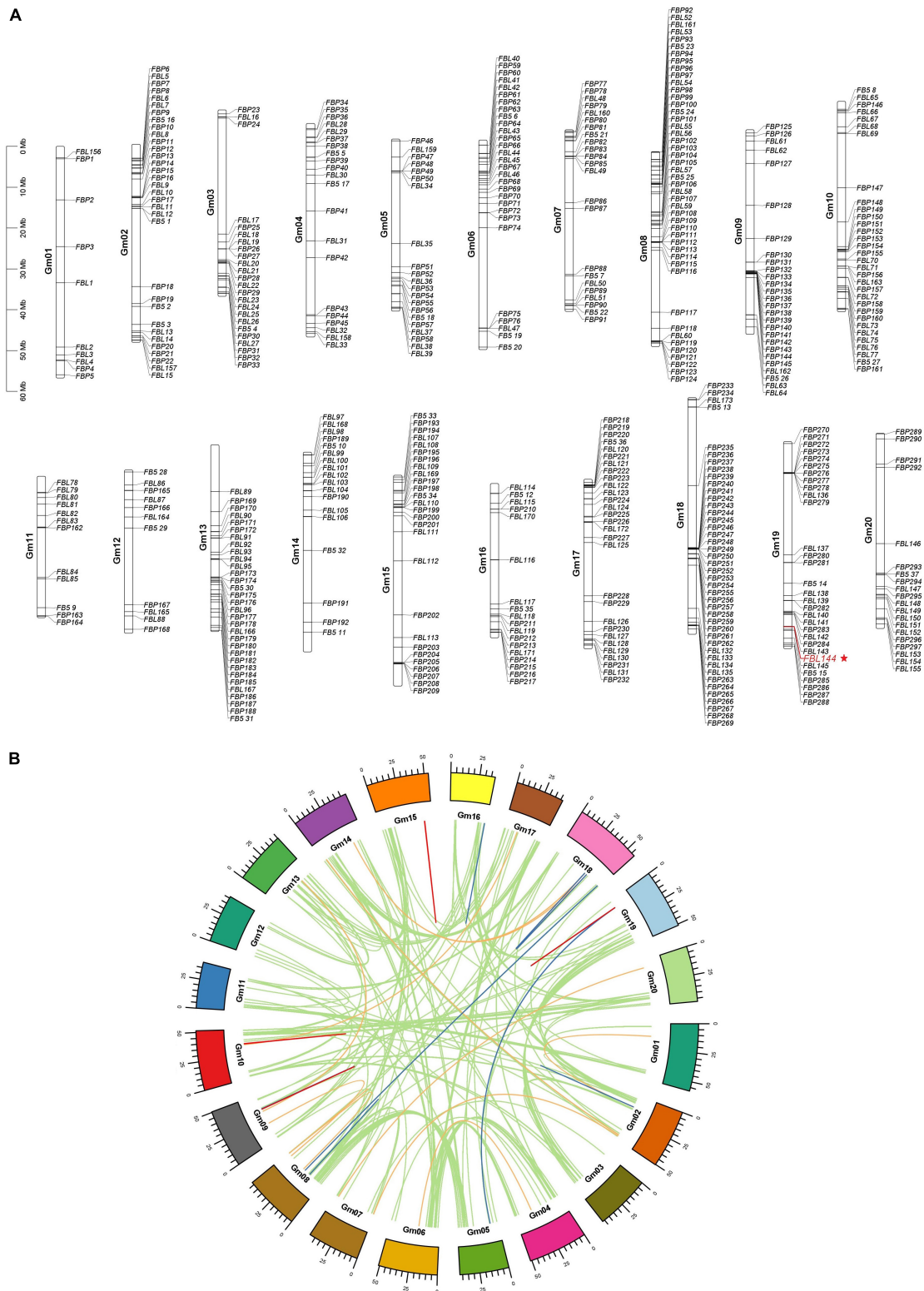


FIGURE 1 | Chromosome distribution and gene duplication analysis of *GmF-box* genes. **(A)** Chromosome distribution of *GmF-box* genes. **(B)** Gene duplication analysis of *GmF-box* genes. The *GmF-box* gene family expansion was primarily caused by WGD/segmental duplication. WGD/segmental duplications gene pairs are shown in green lines; tandem duplication gene pairs are shown in red lines; dispersed duplication gene pairs are shown in orange lines; proximal duplication gene pairs are shown in blue lines.

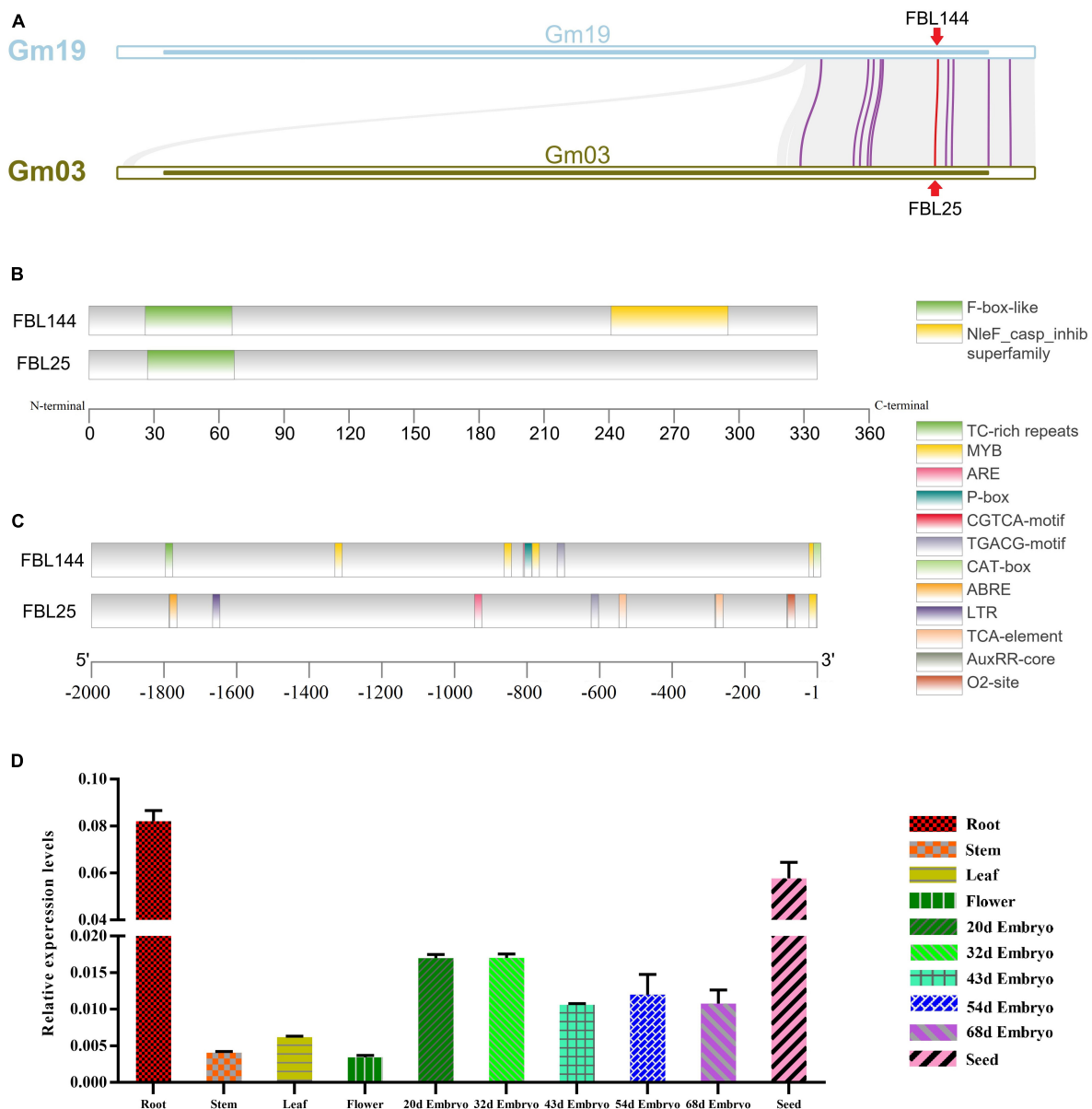


FIGURE 2 | Collinearity analysis and the structural and expression profile analysis of *GmFBL144*. **(A)** Collinearity analysis between chromosome 19 and chromosome 3. **(B)** Gene structure analysis of *GmFBL144* and *GmFBL25*. *GmFBL144* protein shares an F-box domain with *GmFBL25*, and *GmFBL144* has an additional NleF_casp_inhib domain. **(C)** Promoter region analysis of *GmFBL144* and *GmFBL25*. The *GmFBL144* promoter contains four drought response elements (MYB motif) and one stress response element (TC-rich repeats). The *GmFBL25* promoter contains only one drought response element (MYB motif). **(D)** Expression profile analysis of *GmFBL144*. The expression of *GmFBL144* in roots, stem, leaves, flowers, embryos of different stages, and seeds. Relative gene expression was calculated using the $2^{-\Delta Ct}$ method. Data represent means \pm SE of three biological replicates.

GmFBL144 was primarily located on the nucleus, we constructed a soybean cDNA library of a yeast nuclear system. The primary library capacity was approximately 1.04×10^7 , the recombination rate was approximately 100%, and the average insert length was $>1,000$ bp (Figure 7A). The sub-library capacity was approximately 1.44×10^7 , the recombination rate was approximately 100%, and the average insert length was $>1,000$ bp (Figure 7B). After the soybean cDNA library was constructed, we used BD-GmFBL144 as bait for screening interaction proteins

with the Y2H assay. Forty-five positive blue clones were obtained and annotated in **Supplementary Table 3**. Among these clones, small heat shock protein (sHSP) (GenBank: XP_014626034.1), involved in multiple abiotic stress in some plants (Kuang et al., 2017; Guo et al., 2020), was identified for four times, indicating that it might have the strong interaction with GmFBL144.

The Y2H assay results of the one-to-one interaction verification showed that GmFBL144 can interact with GmSHSP (Figure 7C). We further verified the interaction between

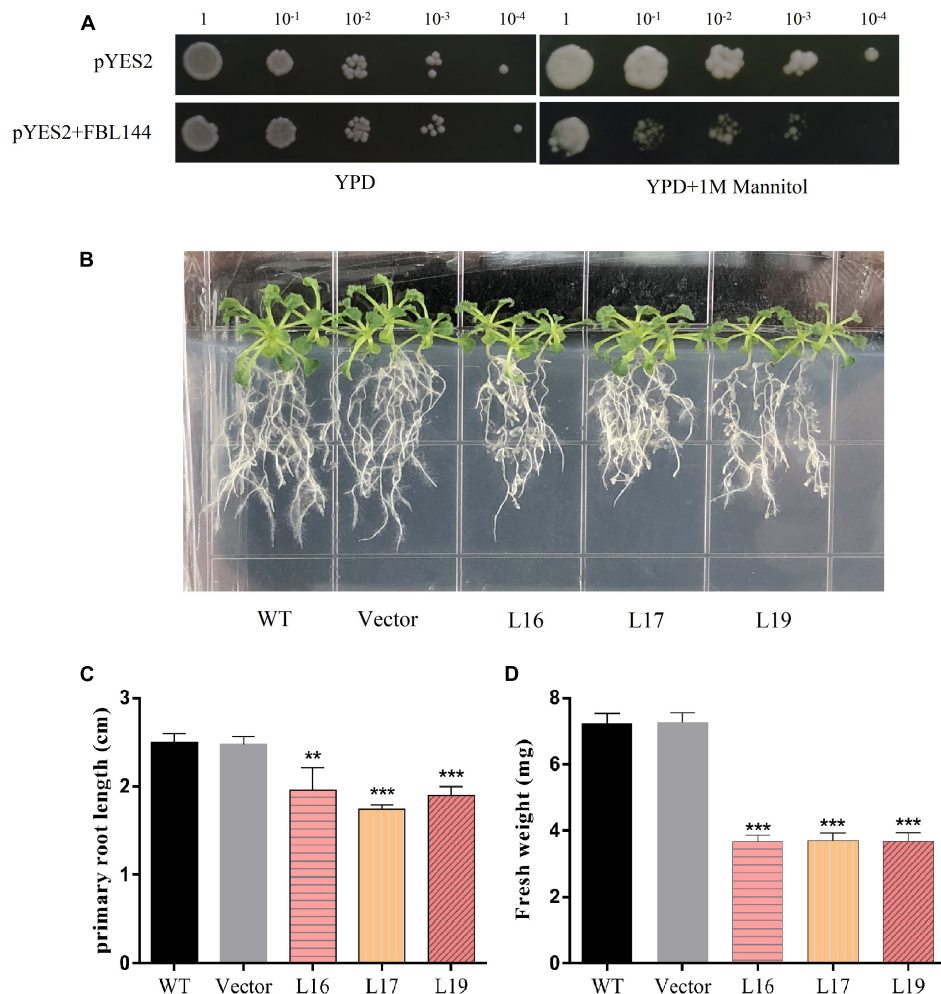


FIGURE 3 | Phenotype of transgenic *Arabidopsis* under simulated drought stress. **(A)** Phenotype of heterologous expression of *GmFBL144* yeast and control under drought stress (1 M mannitol). **(B)** Phenotype of control (wild type and vector) and *GmFBL144*-overexpression lines (L16, L17, and L19) under drought stress (250 mM mannitol). **(C)** Primary root length of control and *GmFBL144*-overexpression lines under drought stress (250 mM mannitol). **(D)** Fresh weight of control and *GmFBL144*-overexpression lines under drought stress (250 mM mannitol). Data represent means \pm SE of three biological replicates. Asterisks indicate significant difference applying ANOVA (* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

GmFBL144 and GmsHSP using BiFC assay. The fusion constructs of FBL144-nYFP and sHSP-cYFP were co-transferred into epidermal *Nicotiana benthamiana* cells for transient assays through *Agrobacterium* infiltration. Yellow fluorescence was observed in cells co-transferred with fusion constructs (Figure 7D), whereas in the negative control, yellow fluorescence was not observed. The results revealed that GmFBL144 can interact with GmsHSP protein.

Small heat shock proteins (sHSPs) are important molecular chaperones, which can prevent damaged protein aggregation caused by stress. sHSP help damaged protein refold to restore biological function by cooperating with other HSPs (HSP100 or HSP70) in the presence of ATP (Lee et al., 2005; Bernfur et al., 2017; Waters and Vierling, 2020). Previous studies have shown sHSP can enhance plant tolerance to external stress (Chauhan et al., 2012; Kuang et al., 2017; Guo et al., 2020). In this study, we identified the function of sHSP through the

yeast transient expression system. The results showed that the growth of overexpression of *GmsHSP* yeast was better than that with vector control under drought stress (Figure 7E). This result shows that GmsHSP is a positive regulator of drought stress. The subcellular localization analysis of GmsHSP found that GmsHSP was mainly localized in nucleus and peroxisome (Supplementary Figure 3).

DISCUSSION

Gene Duplication

In the process of evolution, soybean has experienced two genome duplication or polyploidization events, resulting in a highly duplicated genome with 75% of genes present in the form of paralogous copies (Shoemaker et al., 1996; Gill et al., 2009; Schmutz et al., 2010). In addition, segmental duplication and

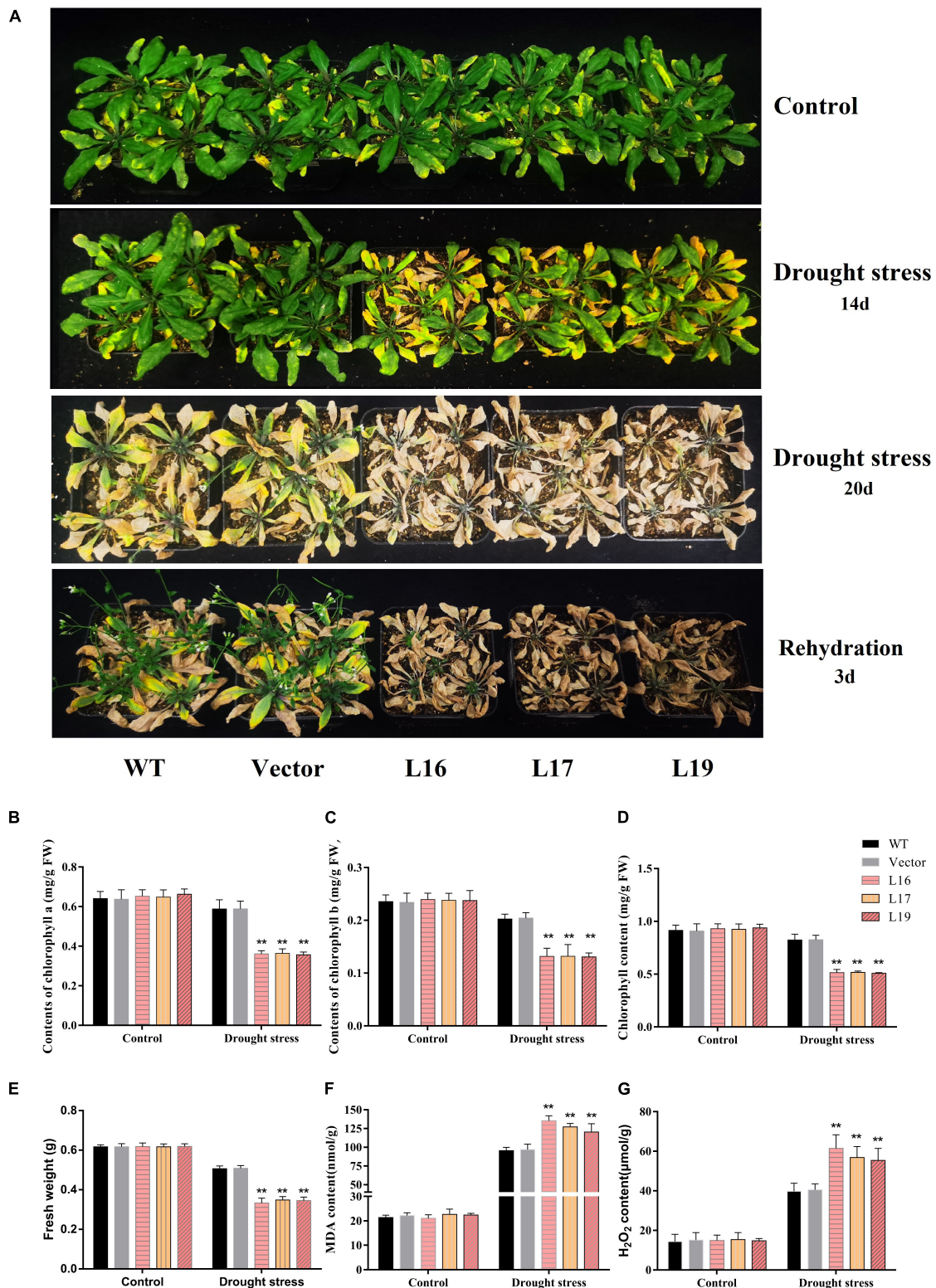
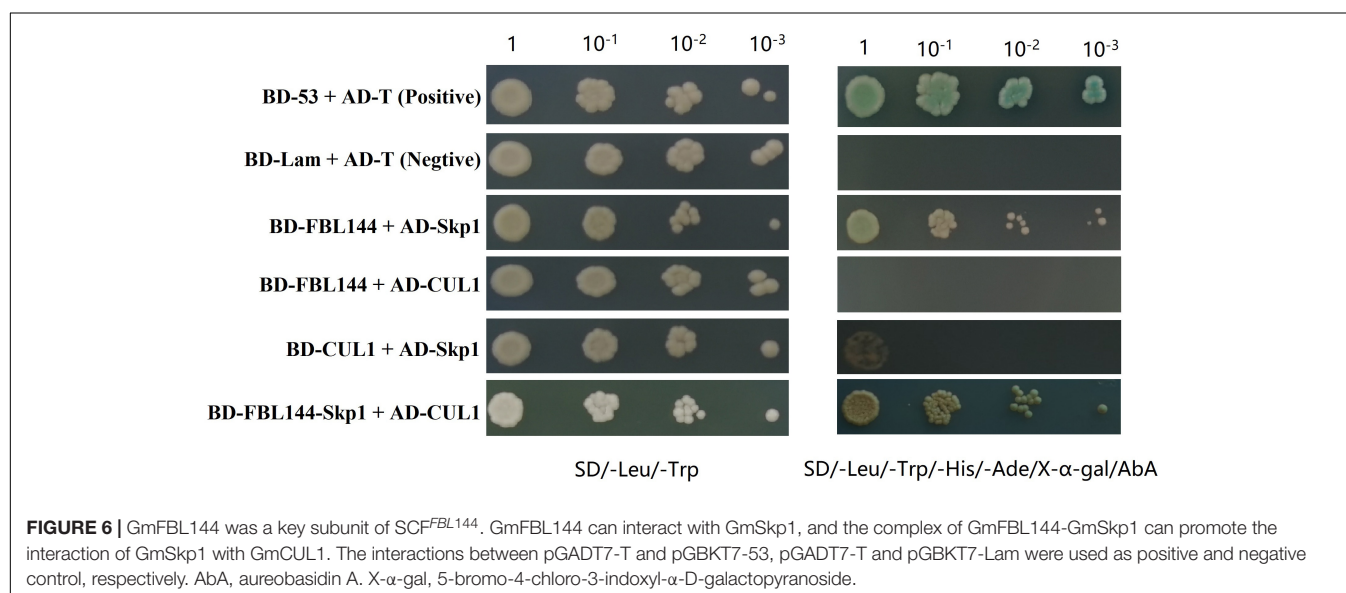
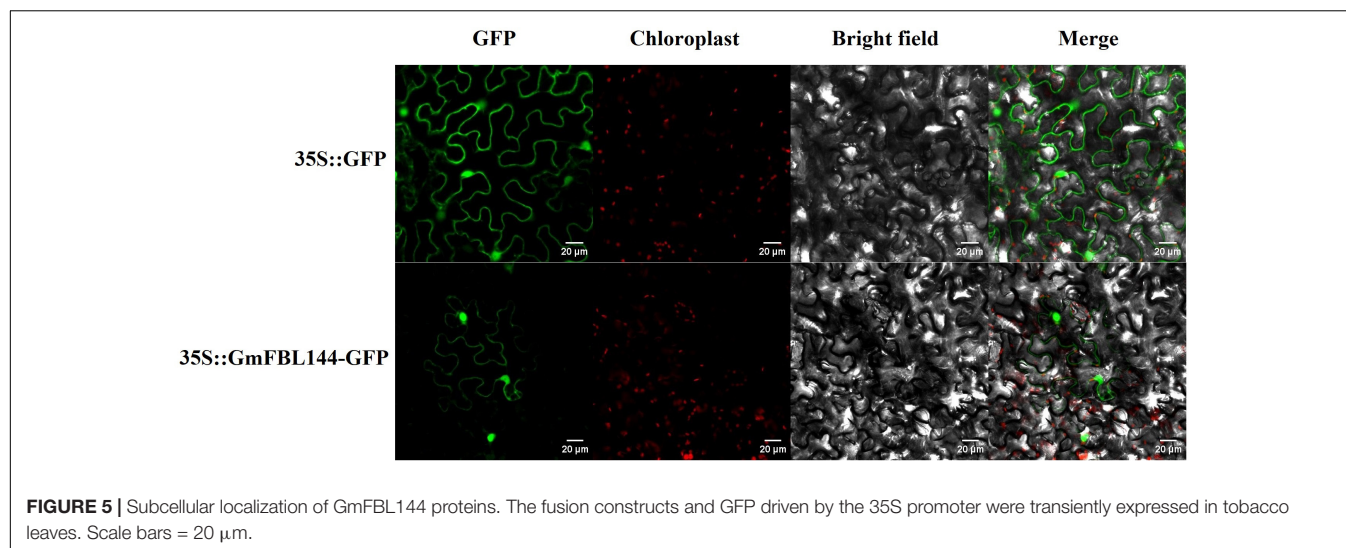


FIGURE 4 | Phenotype and physiological index of transgenic *Arabidopsis* under drought stress. **(A)** Phenotype of control (wild type and vector) and *GmFBL144*-overexpression lines (L16, L17, and L19) under drought stress. **(B–G)** Physiological index of control and *GmFBL144*-overexpression plants under drought stress, including chlorophyll content **(B–D)**, fresh weight **(E)**, MDA content **(F)**, H₂O₂ content **(G)**. Data represent means ± SE of three biological replicates. Asterisks indicate significant difference applying ANOVA (**P* < 0.05; ***P* < 0.01; ****P* < 0.001).



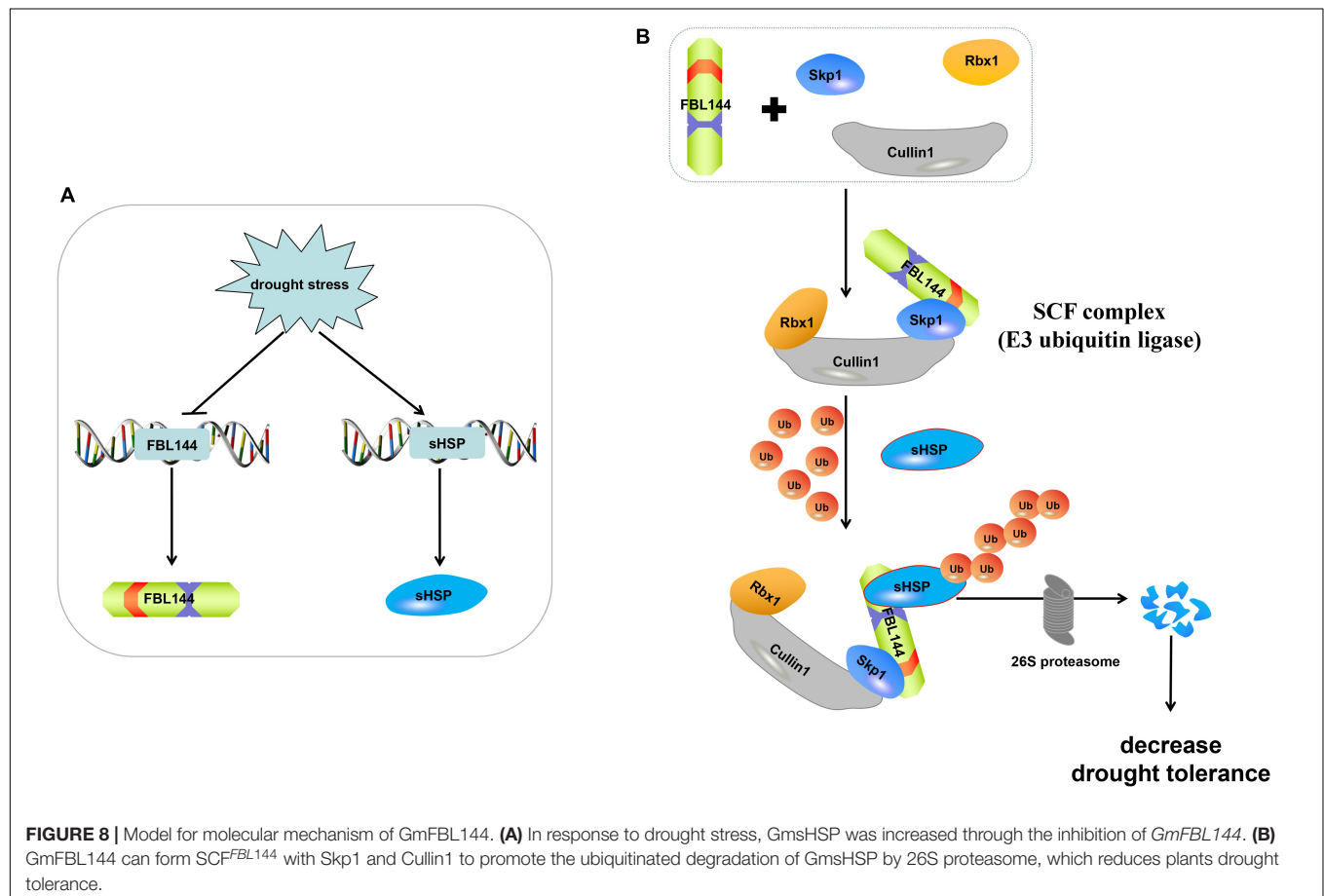
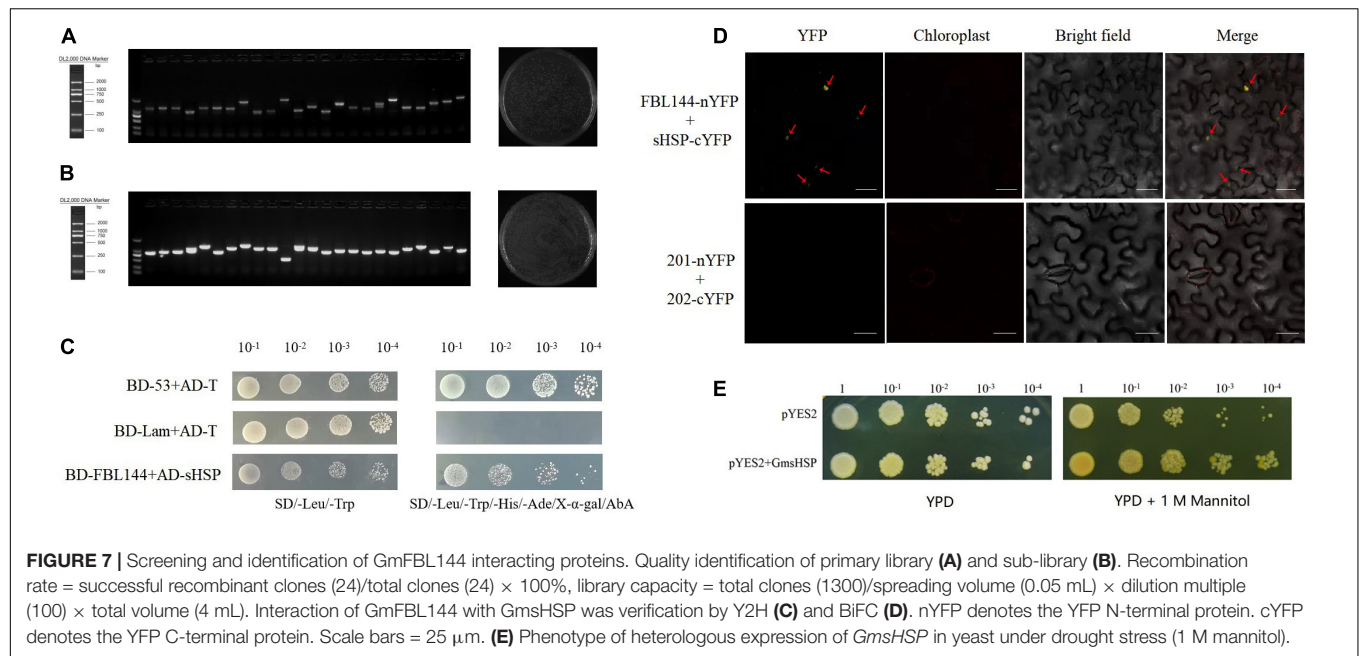
tandem duplication are important reasons for increasing gene copies and genetic diversity (Zhao et al., 2018). Previous studies have shown that *F-box* gene family expansion of rice and chickpea mainly caused by tandem duplication (Jain et al., 2007; Gupta et al., 2015). However, in our study, WGD/segmental duplication was the main expansion form of *F-box* gene in soybean. Similarly, the expansion of the *F-box* gene family in wheat, pear, and cotton was dominated by WGD/segmental duplication (Wang G. et al., 2016; Zhang S. et al., 2019; Li et al., 2020a). These differences may be caused by variations in the definition of tandem duplication. Previous studies of the *F-box* gene family have used different definitions on tandem duplication. In contrast, the tandem duplication definition used by MCSanX is more restrictive than that of previous studies (Wang et al., 2012).

The terminal genes of chromosome 19 and chromosome 3 showed collinearity in addition to some *F-box* genes (Figure 2A), which may have contributed to chromosome rearrangement

after polyploidization. Similar results have also reported in other studies (Schmutz et al., 2010; Zhang et al., 2018). *GmFBL144* and *GmFBL25* have high homology (Supplementary Figure 1 and Figure 2B), which may be the result of evaluation selection of double-copy gene. Evolutionary selection can lead to loss one of the homologous genes or pseudogenization, and the evolution of new function (Moore and Purugganan, 2003; Semon and Wolfe, 2007; Zhao et al., 2018). Future studies will assess whether *GmFBL25* has other functions.

GmFBL144 Is a Key Subunit of the SCF^{FBL144} Complex

The SCF complex is the main form of the F-box protein that performs function. Some F-box proteins perform function in a non-SCF form in yeast and human (Galan et al., 2001; Nelson et al., 2013), but this has not been found in plants.



Interestingly, some F-box proteins can perform functions in both SCF and non-SCF forms in human, for example Fbxo7 (Nelson et al., 2013). Presently, research on plant F-box proteins mainly focuses on their SCF-dependent functions, and there are a large number of F-box proteins in plant. Therefore, it is impossible to rule the existence of the SCF-independent functions of F-box proteins. Previous studies suggested that Cullin1 and Rbx1, Skp1 form a core scaffold (Gagne et al., 2002; Qinxue et al., 2018; Ban and Estelle, 2021). However, in our study, the interaction between Cullin1 and Skp1 was weak, but GmFBL144 can promote the combination of Cullin1 and Skp1 to form SCF^{FBL144}. This result may be related to the binding of GmFBL144 and GmSkp1, which changes the conformation of GmSkp1 resulting in binding-capacity enhancement of GmCullin1 and GmSkp1.

GmFBL144 Enhances the Sensitivity of Plants to Drought Stress

F-box genes play an important role in plant growth, development and stress responses. Recently, considerable research has suggested that F-box genes participate in drought stress responses. For example, *Capsicum annuum* Drought-Induced F-box Protein 1 (CaDIF1) is a positive regulator of drought tolerance (Lim et al., 2019). Under drought stress, *F-BOX OF FLOWERING 2* (*FOF2*) positively regulated ABA-induced stomatal closure, resulting in reduced water loss (Qu et al., 2020). In addition, Li et al. (2020b) found that SITLFP8 (Tubby-like F-box protein 8) can enhance tomato drought tolerance by decreasing water loss *via* changing stomatal density. In general, previous studies have shown that *F-box* genes enhance the drought stress tolerance of plants. However, our study showed that *GmFBL144* enhanced the sensitivity of plants to drought stress, potentially caused by the interaction between GmFBL144 and GmsHSP. sHSP as an important molecular chaperone can maintain protein stability (Morris et al., 2010; Papsdorf and Richter, 2014). They also associate with membranes. Sakthivel et al. (2009) found that HspA can stabilize membrane proteins such as the photosystems and phycobilisomes from oxidative damage. Balogi et al. (2008) found that a mutant Hsp17 (Q16R) with increased thylakoid association can improve the tolerance of UV-B damage in *synechocystis*. In our study, GmsHSP was a positive regulator of drought stress (Figure 7E). When plants were subjected to poor environmental conditions, the homeostasis was maintained through the molecular chaperone and proteolytic systems (Li et al., 2019, 2020d). In our study, *GmFBL144* was down-regulated and *GmsHSP* was up-regulated under drought stress (Supplementary Figure 2). These results indicated that soybean can withstand drought stress by enhancing the chaperone system. Previous studies have found that *F-box* genes positively regulated drought tolerance, likely because F-box proteins maintain intracellular homeostasis *via* the UPS, or increase the content of positive regulatory factors of drought stress. However, in our study, the GmFBL144 may degrade GmsHSP by SCF^{FBL144}, destroying the molecular chaperone system and aggravating the protein homeostasis imbalance that leads to drought sensitivity (Figure 8).

CONCLUSION

In this study, a total of 507 *GmF-box* genes were identified, and classified into 11 subfamilies. The expansion of the *GmF-box* gene family was primarily caused by WGD/segmental duplication. Under drought stress, the expression of *GmFBL144* was down-regulated (Supplementary Figure 2). Overexpression of *GmFBL144* enhanced sensitivity to drought stress. GmFBL144 can form SCF with Skp1 and Cullin1, and interact with GmsHSP. GmFBL144 may promote sHSP ubiquitination through forming SCF^{FBL144}, and then the ubiquitinated sHSP is degraded by 26S proteasome (Figure 8B).

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

KX and YuZ performed the bioinformatics analysis. YaZ, CF, YiZ, FW, XL, HG, WL, YJ, RS, and XF provided assistance on the experiments. HL and YoZ designed the study and revised the manuscript. All authors reviewed and approved the final manuscript.

FUNDING

This study was financially supported by the National Natural Science Foundation of China (32171937 and 32001464), the Natural Science Foundation of Science Technology Department of Jilin Province (20190201259JC), the Hainan Provincial Natural Science Foundation of China (321QN182), the Scientific Research Foundation of Hainan University Program (Y3AZ20024), and the Hainan Yazhou Bay Seed Laboratory (B21Y10905 and B21Y10906).

ACKNOWLEDGMENTS

We thank Bin Liu (Institute of Crop Science, Chinese Academy of Agricultural Sciences) for supplying the plasmid of BiFC. We also thank of Rong Huang for spiritual support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.823529/full#supplementary-material>

REFERENCES

- Ao, K., Tong, M., Li, L., Lüdke, D., Lipka, V., Chen, S., et al. (2020). SCF^{SNIPER7} controls protein turnover of unfoldase CDC48A to promote plant immunity. *New Phytol.* 229, 2795–2811. doi: 10.1111/nph.17071
- Baldoni, E., Genga, A., and Cominelli, E. (2015). Plant MYB transcription factors: their role in drought response mechanisms. *Int. J. Mol. Sci.* 16, 15811–15851. doi: 10.3390/ijms160715811
- Balogi, Z., Cheregi, O., Giese, K. C., Vierling, E., Vass, I., Vigh, L., et al. (2008). A mutant small heat shock protein with increased thylakoid association provides an elevated resistance against UV-B Damage in *Synechocystis*. *J. Biol. Chem.* 283, 22983–22991. doi: 10.1074/jbc.M710400200
- Ban, Z., and Estelle, M. (2021). CUL3 E3 ligases in plant development and environmental response. *Nat. Plants* 7, 6–16. doi: 10.1038/s41477-020-00833-6
- Bernfur, K., Rutsdottir, G., and Emanuelsson, C. (2017). The chloroplast-localized small heat shock protein Hsp21 associates with the thylakoid membranes in heat-stressed plants. *Prot. Sci.* 26, 1773–1784. doi: 10.1002/pro.3213
- Carbonnel, S., Das, D., Varshney, K., Kolodziej, M. C., Villaécija-Aguilar, J. A., and Gutjahr, C. (2020). The karrikin signaling regulator SMAX1 controls Lotus japonicus root and root hair development by suppressing ethylene biosynthesis. *Proc. Natl. Acad. Sci. USA* 117, 21757–21765. doi: 10.1073/pnas.2006111117
- Chauhan, H., Khurana, N., Nijhavan, A., Khurana, J. P., and Khurana, P. (2012). The wheat chloroplastic small heat shock protein (sHSP26) is involved in seed maturation and germination and imparts tolerance to heat stress. *Plant Cell Environ.* 35, 1912–1931. doi: 10.1111/j.1365-3040.2012.02525.x
- Chen, Y., Xu, Y., Luo, W., Li, W., Na, C., and Chong, Z. K. (2013). The F-box protein OsFBK12 targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. *Plant Physiol.* 163, 1673–1685. doi: 10.1104/pp.113.224527
- Clough, S. J., and Bent, A. F. (2010). Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* 16, 735–743. doi: 10.1046/j.1365-313x.1998.00343.x
- Gagne, J. M., Downes, B. P., Shiu, S. H., Durski, A. M., and Vierstra, R. D. (2002). The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 99, 11519–11524. doi: 10.1073/pnas.162339999
- Galan, J. M., Wiederkehr, A., Seol, J. H., Haguenauer-Tsapis, R., Deshaies, R. J., Riezman, H., et al. (2001). Skp1p and the F-Box protein Rcy1p form a non-SCF complex involved in recycling of the SNARE Snc1p in yeast. *Mol. Cell. Biol.* 21, 3105–3117. doi: 10.1128/MCB.21.9.3105-3117.2001
- Gill, N., Findley, S., Walling, J. G., Hans, C., Ma, J., Doyle, J., et al. (2009). Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiol.* 151, 1167–1174. doi: 10.1104/pp.109.137935
- Gong, Z., Xiong, L., Shi, H., Yang, S., and Zhu, J. K. (2020). Plant abiotic stress response and nutrient use efficiency. *Sci. China Life Sci.* 63, 635–674. doi: 10.1007/s11427-020-1683-x
- Guo, L. M., Li, J., He, J., Liu, H., and Zhang, H. M. (2020). A class I cytosolic HSP20 of rice enhances heat and salt tolerance in different organisms. *Sci. Rep.* 10:1383. doi: 10.1038/s41598-020-58395-8
- Gupta, S., Garg, V., Kant, C., and Bhatia, S. (2015). Genome-wide survey and expression analysis of F-box genes in chickpea. *BMC Genom.* 16:67. doi: 10.1186/s12864-015-1293-y
- Jain, M., Nijhawan, A., Arora, R., Agarwal, P., Ray, S., Sharma, P., et al. (2007). F-box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol.* 143, 1467–1483. doi: 10.1104/pp.106.091900
- Jie, L., Qfla, E., Shi, L., Xuan, J., Hui, L., Tao, X. B., et al. (2020). The F-box protein EST1 modulates salt tolerance in Arabidopsis by regulating plasma membrane Na⁺/H⁺ antiport activity. *J. Plant Physiol.* 251, 153217. doi: 10.1016/j.jplph.2020.153217
- Kipreos, E. T., and Pagano, M. (2000). The F-box protein family. *Genome Biol.* 1, 1–7.
- Kuang, J., Liu, J., Mei, J., Wang, C., Hu, H., Zhang, Y., et al. (2017). A Class II small heat shock protein OsHsp18.0 plays positive roles in both biotic and abiotic defense responses in rice. *Sci. Rep.* 7:11333. doi: 10.1038/s41598-017-11882-x
- Lee, U., Wie, C., Escobar, M., and Williams, B. (2005). Genetic analysis reveals domain interactions of Arabidopsis Hsp100/ClpB and cooperation with the small heat shock protein chaperone system. *Plant Cell* 17, 559–571. doi: 10.1105/tpc.104.027540
- Lescot, M., Dehais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., et al. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 30, 325–327. doi: 10.1093/nar/30.1.325
- Li, C., Li, K., Liu, X., Ruan, H., Zheng, M., Yu, Z., et al. (2021). Transcription factor GmWRKY46 enhanced phosphate starvation tolerance and root development in transgenic plants. *Front. Plant Sci.* 12:700651. doi: 10.3389/fpls.2021.700651
- Li, H., Wei, C., Meng, Y., Fan, R., Zhao, W., Wang, X., et al. (2020a). Identification and expression analysis of some wheat F-box subfamilies during plant development and infection by *Puccinia triticina*. *Plant Physiol. Bioch.* 155, 535–548. doi: 10.1016/j.plaphy.2020.06.040
- Li, S., Zhang, J., Liu, L., Wang, Z., Li, Y., Guo, L., et al. (2020b). SITLFP8 reduces water loss to improve water-use efficiency by modulating cell size and stomatal density via endoreduplication. *Plant Cell Environ.* 43, 2666–2679. doi: 10.1111/pce.13867
- Li, X., Chen, P., Xie, Y., Yan, Y., and Guan, Q. (2020c). Apple SERRATE negatively mediates drought resistance by regulating MdMYB88 and MdMYB124 and microRNA biogenesis. *Hortic. Res.* 7:98. doi: 10.1038/s41438-020-0320-6
- Li, X., Liu, Q., Feng, H., Deng, J., Zhang, R., Wen, J., et al. (2020d). Dehydrin MtCAS31 promotes autophagic degradation under drought stress. *Autophagy* 16, 862–877. doi: 10.1080/15548627.2019.1643656
- Li, Y., Xue, Y., Xu, X., Wang, G., Liu, Y., Wu, H., et al. (2019). A mitochondrial FUNDC1/HSC70 interaction organizes the proteostatic stress response at the risk of cell morbidity. *EMBO J.* 38:e98786. doi: 10.15252/emboj.201798786
- Lim, J., Lim, C. W., and Lee, S. C. (2019). Functional analysis of pepper F-box protein CaDIF1 and its interacting partner CaDIS1: modulation of ABA signaling and drought stress response. *Front. Plant Sci.* 10:1365. doi: 10.3389/fpls.2019.01365
- Lu, Q., Tang, X., Tian, G., Wang, F., Liu, K., Nguyen, V., et al. (2009). Arabidopsis homolog of the yeast TREX-2 mRNA export complex: components and anchoring nucleoporin. *Plant J.* 61, 259–270. doi: 10.1111/j.1365-313X.2009.04048.x
- Moore, R. C., and Purugganan, M. D. (2003). The early stages of duplicate gene evolution. *Proc. Natl. Acad. Sci. USA* 100, 15682–71568.
- Morreale, F. E., and Walden, H. (2016). Types of ubiquitin ligases. *Cell* 165, 248–248.
- Morris, A. M., Treweek, T. M., Aquilina, J. A., Carver, J. A., and Walker, M. J. (2010). Glutamic acid residues in the C-terminal extension of small heat shock protein 25 are critical for structural and functional integrity. *FEBS J.* 275, 5885–5898. doi: 10.1111/j.1742-4658.2008.06719.x
- Nelson, D. E., Randle, S. J., and Laman, H. (2013). Beyond ubiquitination: the atypical functions of Fbxo7 and other F-box proteins. *Open Biol.* 3:130131. doi: 10.1098/rsob.130131
- Papsdorf, K., and Richter, K. (2014). Protein folding, misfolding and quality control: the role of molecular chaperones. *Essays Biochem.* 56, 53–68. doi: 10.1042/bse0560053
- Qinxue, L., Wenqiang, W., Wenlong, W., Guangqiang, Z., Yang, L., Yong, W., et al. (2018). Wheat F-Box protein gene TaFBA1 is involved in plant tolerance to heat stress. *Front. Plant Sci.* 9:521. doi: 10.3389/fpls.2018.00521
- Qu, L., Sun, M., Li, X., He, R., and Zhao, X. (2020). The Arabidopsis F-box protein FOF2 regulates ABA-mediated seed germination and drought tolerance. *Plant Sci.* 301:110643. doi: 10.1016/j.plantsci.2020.110643
- Ritchie, S. W., Nguyen, H. T., and Holaday, A. S. (1990). Leaf water content and gas-exchange parameters of two wheat genotypes differing in drought resistance. *Crop. Sci.* 30, 105–111.
- Sadanandom, A., Bailey, M., Ewan, R., Lee, J., and Nelis, S. (2012). The ubiquitin-proteasome system: central modifier of plant signalling. *New Phytol.* 196, 13–28. doi: 10.1111/j.1469-8137.2012.04266.x
- Sakthivel, K., Watanabe, T., and Nakamoto, H. (2009). A small heat-shock protein confers stress tolerance and stabilizes thylakoid membrane proteins in *cyanobacteria* under oxidative stress. *Arch. Microbiol.* 191, 319–328. doi: 10.1007/s00203-009-0457-z
- 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

- Semon, M., and Wolfe, K. H. (2007). Consequences of genome duplication. *Curr. Opin. Genet. Dev.* 17, 505–512.
- Shoemaker, R. C., Polzin, K., Labate, J., Specht, J., Brummer, E. C., Olson, T., et al. (1996). Genome duplication in soybean (*Glycine subgenus soja*). *Genetics* 144, 329–329. doi: 10.1093/genetics/144.1.329
- Venkatesh, J., Kang, M. Y., Liu, L., Kwon, J. K., and Kang, B. C. (2020). F-Box family genes, *LTSF1* and *LTSF2*, regulate low-temperature stress tolerance in pepper (*Capsicum chinense*). *Plants* 9:1186. doi: 10.3390/plants9091186
- Wang, G., Yin, H., Qiao, X., Tan, X., Gu, C., Wang, B., et al. (2016). F-box genes: Genome-wide expansion, evolution and their contribution to pollen growth in pear (*Pyrus bretschneideri*). *Plant Sci.* 253, 164–175. doi: 10.1016/j.plantsci.2016.09.009
- Wang, X., Wang, H., Liu, S., Ferjani, A., Li, J., Yan, J., et al. (2016). Genetic variation in *ZmVPP1* contributes to drought tolerance in maize seedlings. *Nat. Genet.* 48, 1233–1241. doi: 10.1038/ng.3636
- Wang, Y., Li, J., and Paterson, A. H. (2013). MCSanX-transposed: detecting transposed gene duplications based on multiple colinearity scans. *Bioinformatics* 29, 1458–1460. doi: 10.1093/bioinformatics/btt150
- Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40:e49. doi: 10.1093/nar/gkr1293
- Waters, E. R., and Vierling, E. (2020). Plant small heat shock proteins – evolutionary and functional diversity. *New Phytol.* 227, 24–37. doi: 10.1111/nph.16536
- Woo, H. R., Chung, K. M., Park, J. H., Oh, S. A., Ahn, T., Hong, S. H., et al. (2001). *ORE9*, an F-Box protein that regulates leaf senescence in Arabidopsis. *Plant Cell* 13, 1779–1790. doi: 10.1105/tpc.010061
- Xia, F. N., Zeng, B., Liu, H. S., Qi, H., Xie, L. J., Yu, L. J., et al. (2020). SINAT E3 ubiquitin ligases mediate FREE1 and VPS23A degradation to modulate abscisic acid signaling. *Plant Cell* 32, 3290–3310. doi: 10.1105/tpc.20.00267
- Zafari, M., Ebadi, A., Jahanbakhsh, S., and Sedghi, M. (2020). Safflower (*Carthamus tinctorius*) biochemical properties, yield, and oil content affected by 24-epibrassinosteroid and genotype under drought stress. *J. Agric. Food Chem.* 68, 6040–6047. doi: 10.1021/acs.jafc.9b06860
- Zhang, S., Tian, Z., Li, H., Guo, Y., Zhang, Y., Roberts, J. A., et al. (2019). Genome-wide analysis and characterization of F-box gene family in *Gossypium hirsutum* L. *BMC Genom.* 20:1–16.
- Zhang, Y., Jian, L., Chai, J., and Da, X. (2016). Mitogen-activated protein kinase 6 mediates nuclear translocation of ORE3 to promote ORE9 gene expression in methyl jasmonate-induced leaf senescence. *J. Exp. Bot.* 67, 83–94. doi: 10.1093/jxb/erv438
- Zhang, Y., Zhang, J., Guo, J., Zhou, F., Singh, S., Xu, X., et al. (2019). F-box protein RAE1 regulates the stability of the aluminum-resistance transcription factor STOP1 in Arabidopsis. *Proc. Natl. Acad. Sci. USA* 116, 319–327. doi: 10.1073/pnas.1814426116
- Zhang, Z., Zhao, Y., Feng, X., Luo, Z., Kong, S., Zhang, C., et al. (2018). Genomic, molecular evolution, and expression analysis of NOX genes in soybean (*Glycine max*). *Genomics* 111, 619–628. doi: 10.1016/j.ygeno.2018.03.018
- Zhao, M., Zhang, B., Ma, J., and Lisch, D. (2018). Genome-wide estimation of evolutionary distance and phylogenetic analysis of homologous genes. *Bio-Protocol* 8:e3097. doi: 10.21769/BioProtoc.3097
- Zheng, N., and Shabek, N. (2013). Ubiquitin Ligases: structure, function, and regulation. *Annu. Rev. Biochem.* 86, 129–157. doi: 10.1146/annurev-biochem-060815-014922
- Zheng, N., Zhou, Q., Wang, Z., and Wei, W. (2016). Recent advances in SCF ubiquitin ligase complex: Clinical implications. *Biochim. Biophys. Acta.* 1866, 12–22. doi: 10.1016/j.bbcan.2016.05.001
- Zhou, Y., Liu, W., Li, X., Sun, D., Xu, K., Feng, C., et al. (2020). Integration of sRNA, degradome, transcriptome analysis and functional investigation reveals gma-miR398c negatively regulates drought tolerance via *GmCSDs* and *GmCCS* in transgenic Arabidopsis and soybean. *BMC Plant Biol.* 20:190. doi: 10.1186/s12870-020-02370-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 Xu, Zhao, Feng, Zhang, Wang, Li, Gao, Liu, Jing, Saxena, Feng, Zhou 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.



Progress in Soybean Genetic Transformation Over the Last Decade

Hu Xu¹, Yong Guo², Lijuan Qiu^{2*} and Yidong Ran^{1*}

¹ Tianjin Genovo Biotechnology Co., Ltd., Tianjin, China, ² Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Madan K. Bhattacharyya,
Iowa State University, United States

Reviewed by:

Shuo Li,
Shandong University, China
Milind B. Ratnaparkhe,
ICAR-Indian Institute of Soybean
Research, India

*Correspondence:

Lijuan Qiu
qiu lijuan@caas.cn
Yidong Ran
yidongran@genovo.org

Specialty section:

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

Received: 20 March 2022

Accepted: 11 May 2022

Published: 09 June 2022

Citation:

Xu H, Guo Y, Qiu L and Ran Y
(2022) Progress in Soybean Genetic
Transformation Over the Last Decade.
Front. Plant Sci. 13:900318.
doi: 10.3389/fpls.2022.900318

Soybean is one of the important food, feed, and biofuel crops in the world. Soybean genome modification by genetic transformation has been carried out for trait improvement for more than 4 decades. However, compared to other major crops such as rice, soybean is still recalcitrant to genetic transformation, and transgenic soybean production has been hampered by limitations such as low transformation efficiency and genotype specificity, and prolonged and tedious protocols. The primary goal in soybean transformation over the last decade is to achieve high efficiency and genotype flexibility. Soybean transformation has been improved by modifying tissue culture conditions such as selection of explant types, adjustment of culture medium components and choice of selection reagents, as well as better understanding the transformation mechanisms of specific approaches such as *Agrobacterium* infection. Transgenesis-based breeding of soybean varieties with new traits is now possible by development of improved protocols. In this review, we summarize the developments in soybean genetic transformation to date, especially focusing on the progress made using *Agrobacterium*-mediated methods and biolistic methods over the past decade. We also discuss current challenges and future directions.

Keywords: soybean transformation, transformation efficiency, genotype, *Agrobacterium*, biolistic method, genome editing

INTRODUCTION

Soybean [*Glycine max* (L.) Merrill] is a legume crop belonging to the family of Leguminosae, a subfamily of Papilionoideae. Soybean is grown worldwide and is one of the most important crop plants for its high seed oil and protein content, and for its capability to fix atmospheric nitrogen by symbioses with soil-borne microorganisms. Recent studies on high-quality reference genome sequencing of a United States variety, Williams82 (Schmutz et al., 2010), a Japanese variety, Enrei (Shimomura et al., 2015), a Chinese cultivar, Zhonghuang13, and a wild soybean, W05 (Shen et al., 2018; Xie et al., 2019) have estimated that there exist a total of 46,430 protein-coding genes in soybean, 70% more than that in *Arabidopsis*. Soybean is an ancient polyploidy (palaeopolyploid) plant with a highly duplicated genome. Nearly 75% of the genes are present in multiple copies, representing a threefold redundancy due to its long evolutionary history (Schmutz et al., 2010). Some repetitive sequence families may be species-specific (Morgante et al., 1997). Several other databases have been developed, including an expressed sequence tag (EST) database, full-length cDNAs and cDNA microarrays (Stacey et al., 2004; Umezawa et al., 2008), and a haplotype map (GmHapMap) (Torkamaneh et al., 2019). These resources provide a wide range of opportunities

for soybean improvement by marker-assisted breeding and with transgenic and genome editing approaches, and for understanding gene function through various forward and reverse genetic approaches. Most of these approaches are reliant on high-throughput transformation systems.

Genetic transformations allow for various genes of interest to be introduced and expressed in cells of living organisms, which can also overcome barriers of sexual incompatibility. Soybean genetic transformation was originally developed in late 1980s. The first fertile transgenic soybeans were produced by either regeneration of cotyledonary nodes infected with *Agrobacterium tumefaciens* (Hinchee et al., 1988) or by particle bombardment using meristems of immature soybean seeds (McCabe et al., 1988). The development of soybean transgenic methods before 2013 has previously been extensively reviewed (Homrich et al., 2012; Yamada et al., 2012; Lee et al., 2013; Mariashibu et al., 2013). Soybean improvements using these transformation methods have been continued over the last 30 years. Since the first transgenic herbicide-resistant soybean product was commercialized in the mid 1990s, soybean has become one of the most important crops improved using modern biotechnology and one of the major commercially grown transgenic plants around the world. Genetically modified (GM) soybean, especially the GM Roundup Ready soybean resistant to glyphosate herbicides, has been grown in many countries including the United States, Argentina, and Brazil (Pagano and Miransari, 2016), which has made it a leading biotech crop. This soybean variety allows for growers to spray herbicides to kill any weeds in the field while not killing the soybean crop¹. It was reported that about 105 million hectares of GM soybean was grown in 2017, and that about 272 million metric tons of seeds were produced, which accounted for 80% of all soybean production in the world (Voora et al., 2020). Genetic engineering has been conducted to enhance the protein quality of soybean by altering biosynthetic feedback pathways that increase lysine and sulfur-containing amino acids (Falco et al., 1995). Many types of GM soybeans have improved traits such as increased oleic acid content, decreased linolenic acid content, delayed flowering time, modified plant architecture and increased yield (Yamada et al., 2012). With increasing soybean demands around the world, especially from China, developing GM soybean varieties with high quality and yield is the main task for soybean researchers and breeders. Recently, genome editing (GE) technologies, especially the clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) technology, have been used for studying soybean genetics and commercial trait development (reviewed in Xu et al., 2020). However, using genome editing technologies on plants has been heavily dependent on efficient transformation systems and regeneration of plants containing edited events (Ran et al., 2017; Gao, 2021). Therefore, an efficient and genotype-flexible transformation system is key to realizing soybean improvement using these new technologies. Unfortunately, soybean remains recalcitrant to routine transformations compared to other major cultivated crops such as rice (Chen et al., 2018b). Low transformation frequency and genotype inflexibility are major

hurdles that limit soybean transgenesis and breeding. In this review, we will summarize the major achievements that have been made in this field since 2013, and describe current best methods used for achieving stable and transient transformations in soybean. We also describe the remaining challenges that need to be addressed.

CURRENT TRANSFORMATION METHODS DEVELOPED FOR SOYBEAN

Various transformation methods have been developed for soybean. Here, we will summarize each transformation method and its ability to produce either stable transgenic plants or transient events used for soybean research (Figure 1).

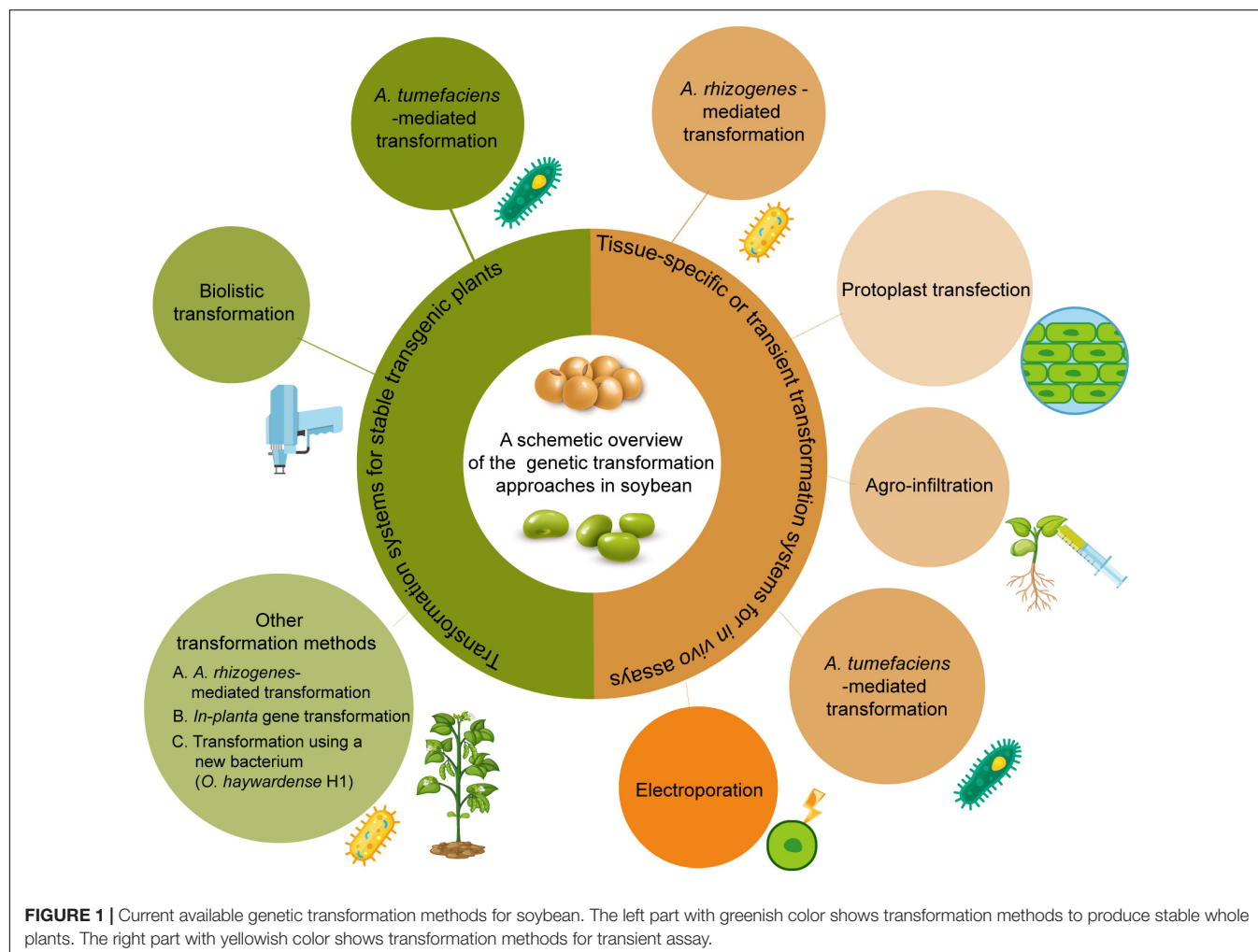
Tissue-Specific or Transient Transformation Systems for *in vivo* Assays

Transient assays are used for a variety of studies including the functional genomics of *in vivo* gene expression and subcellular gene localization, and determination of genome editing efficiency. For soybean, *Agrobacterium rhizogenes*-mediated transformation, protoplast transfection, Agro-infiltration, and electroporation have been developed. *Agrobacterium* transformation and protoplast transfection are frequently performed for transient assays.

Agrobacterium rhizogenes-Mediated Transformation

Agrobacterium rhizogenes-mediated transformation leads to development of a hairy-root phenotype. This method relies on co-transfer of T-DNAs from the Ri plasmid and a binary vector containing a gene of interest into the plant genome (Christey, 2001; Broothaerts et al., 2005). Large numbers of transgenic hairy-roots can be obtained in the absence of exogenous plant growth regulators (Collier et al., 2005), and each represents an independent transformation event (Kereszt et al., 2007). The relatively short timeframe (approximately 6–8 weeks) for recovering transformants is a major advantage for screening genes and promoters or expressing foreign genes in a stably transformed plant as a bioreactor (Cho et al., 2000; Bahramnejad et al., 2019). This method is also used for studying functional genomics in soybean roots. This approach has been used to characterize promoters (Hernandez-Garcia et al., 2009; Li et al., 2014), propagation of nematodes (Cho et al., 2000), symbiotic interactions (Hayashi et al., 2012), pathogenic interactions (Li et al., 2010), gene silencing by RNA interference (RNAi) (Subramanian et al., 2004), and recently for measuring genome editing activity (Du et al., 2016; Cheng et al., 2021). Recently, a reporter gene *AtMyb75*, encoding an R2R3 type MYB transcription factor, was ectopically expressed in hairy roots-mediated by *A. rhizogenes* and induced purple/red colored anthocyanin accumulation in soybean hairy roots. This is a convenient, non-destructive, low cost, directly visual selection of transgenic hairy roots (Fan et al., 2020). Several efficient transformation protocols have been developed for studying

¹<https://www.sourcewatch.org/index.php?title=SourceWatch>



functional genomics and root biology (Kereszt et al., 2007; Kuma et al., 2015; Chen et al., 2018a,c; Fan et al., 2020; Song et al., 2021).

Protoplast Transfection

The first genetic transformation of soybean protoplasts was achieved by electroporation by Lin et al. (1987). Dhir et al. (1992) was the first to report the transformation of immature cotyledon-derived protoplasts and regeneration of transgenic plants from calli derived from electroporation-transfected protoplasts. Protoplasts could be a good explant for transformation if an efficient regeneration system is established, especially since a large number of protoplasts can be transfected at a time and many forms of genetic materials such as DNA, RNA, and protein can be delivered. Unfortunately, protoplast transfection has not yet been conducted for soybean transgenic plant production. The main challenge is achieving protoplast regeneration, which has yet to be reported in soybean. Protoplast-based transfection has been mainly conducted to evaluate gene functions (Yi et al., 2010; Faria et al., 2011; Kidokoro et al., 2015; Xiong et al., 2019), screen promoters (Sultana et al., 2019), and validate vectors for GE (Sun et al., 2015; Demorest et al., 2016; Do et al., 2019; Patil et al., 2022). Recently, Wu and Hanzawa (2018)

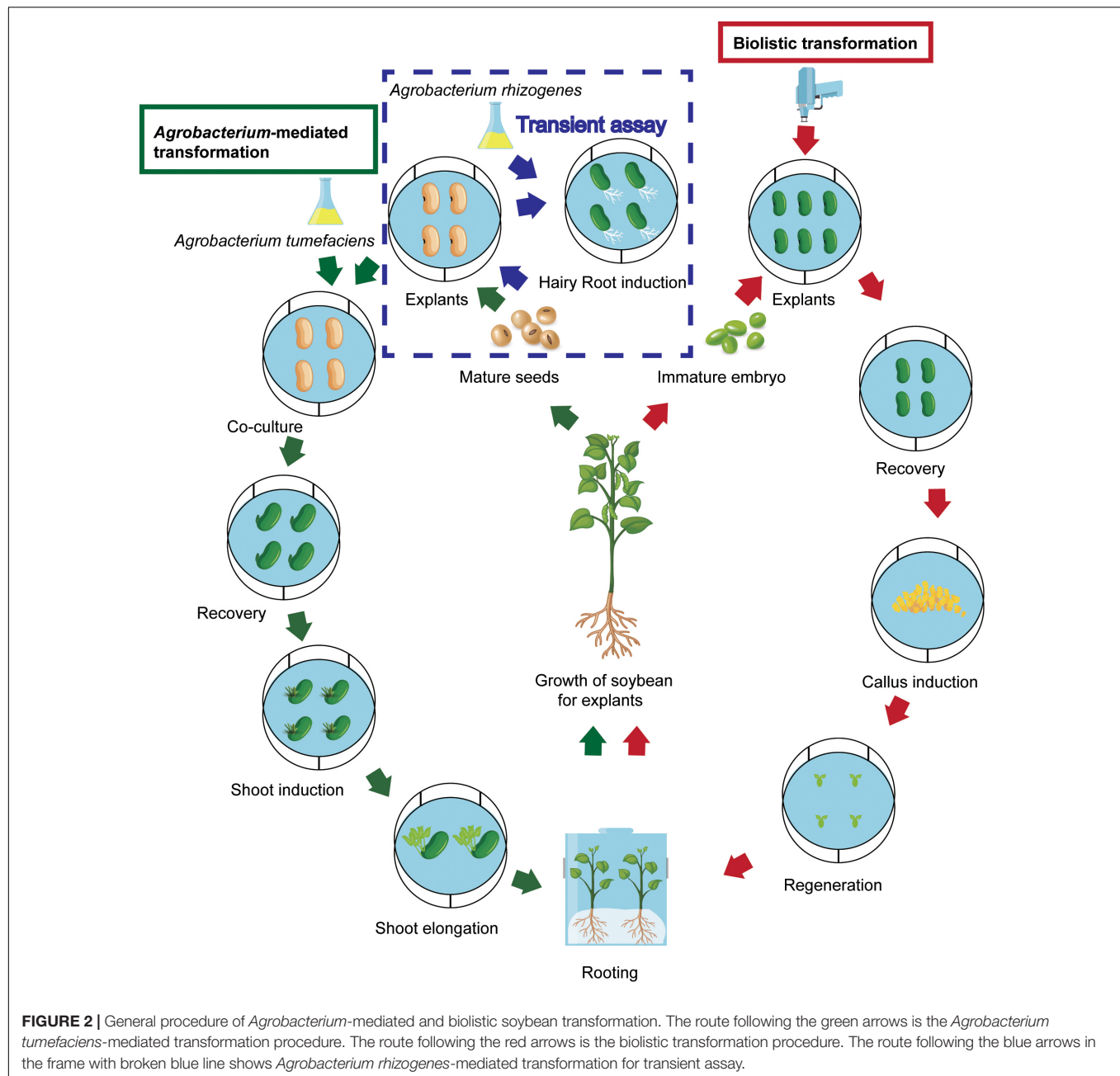
developed a method to isolate protoplasts from leaves of soybean seedlings and established a PEG-mediated transfection method that can achieve high transfection efficiency compared to other transient assays.

Agro-Infiltration

Agrobacteria can be infiltrated into the intercellular space of plant tissues to permit the delivery of genes from different organisms into plant genomes (Grimsley et al., 1986). Ever since this method was successfully established for soybean (King et al., 2015), it has been used for virus-induced gene silencing (VIGS) (Kim et al., 2016) and expression of hairpin RNA for RNAi against two-spot spider mites (Dubey et al., 2017).

Agrobacterium tumefaciens-Mediated Transformation

Except for stable transformation, *A. tumefaciens* is used to carry out transformation for soybean transient assay. Kun et al. (2017) established an *Agrobacterium*-mediated transient system using calli induced from hypocotyl explants. It has been successfully used in many specific assays including Western blot and Co-IP assay for protein analysis. The system is genotype-flexible



and cost-saving. However, it takes a couple of months to complete the assay.

Electroporation

Electroporation is a technique that utilizes a high intensity electric pulse to create transient pores in the cell membrane, thereby facilitating the uptake of macromolecules such as DNA. Christou et al. (1988) conducted electroporation to deliver constructs into soybean calli and showed stable integration of genes but did not succeed in regenerating plants. Later, Chowrira et al. (1995) reported on electroporation of intact nodal

meristems which avoided the soybean tissue culture process completely, but no transgenic plants have been recovered.

Transformation Systems for Stable Transgenic Plants

Agrobacterium-mediated transformation and biolistic methods, and *in planta* transformation and protoplast transfection methods have been applied for generation of transgenic soybean plants. Among these methods, the *A. tumefaciens*-mediated and biolistic methods are the two major platforms for stable soybean transformation. The general transformation procedure of both methods is shown in **Figure 2**. The other methods mentioned

above are used less because of relatively low efficiency and the specific technique and equipment required in these methods.

A. *tumefaciens*-Mediated Transformation

A. tumefaciens-mediated transformation of soybean was first initiated using cotyledonary nodes by Hinchee et al. (1988). Since the system was established based on regeneration of mature or immature seed explants, the simplicity and relatively high TF of the method have made it a favorite method for soybean. Relatively high efficient *Agrobacterium*-mediated transformation protocol has been gradually developed through improving factors such as using an appropriate *Agrobacterium* strain, a good explant, culture media with adequate antioxidant chemicals and combinations of appropriate plant growth regulators for a specific soybean genotype (reviewed in Yamada et al., 2012; Lee et al., 2013; Li et al., 2017; **Table 1**). Key elements of the progress are summarized in a later section. Main advantages of *Agrobacterium* transformation include relatively high ratio of single-copy gene insertion, relative simplicity of the transformation procedure, and low cost (Hwang et al., 2017). However, there is a limitation in delivery of genetic material. It delivers DNA plasmids but cannot deliver DNA fragments, RNAs, or proteins.

Biolistic Transformation

Biolistic transformation, known as gene gun or particle bombardment, delivers small tungsten or gold particles coated with desired genes to target plant cells (Christou et al., 1988). Since an electrical-discharge gene gun was first used in soybean to regenerate a fertile transgenic plant (McCabe et al., 1988), gene delivery to meristematic soybean cells by particle bombardment has been considered to be more genotype-flexible for transfer of foreign DNA into soybean (Homrich et al., 2012). Recently, embryogenic callus based biolistic method becomes more popular due to its relatively higher efficiency compared to other explants and its directly delivering way which meets the need for genome editing using RNA and RNPs editing reagents for recovery of DNA-free edited events. In comparison to the *A. tumefaciens*-mediated method, the biolistic method offers benefits with their capacity to transform organelles and deliver RNA, proteins, nanoparticles, dyes, and complexes to cells (Klein et al., 1987; Liang et al., 2017). The drawback is mainly high transgene copy and relatively high cost, and its application has been restricted in limited soybean genotypes because of unavailable meristematic explants. Compared to plasmid bombardment, utilization of specific constructs including linear minimal expression cassettes (MECs) in biolistic transformation enables the production of plants carrying much simpler patterns of transgene integration, which has been confirmed in plants such as wheat (Ismagul et al., 2018). The major progress in soybean biolistic transformation is presented in a later section and summarized in **Table 1**.

Other Stable Transformation Methods

A. *rhizogenes*-Mediated Transformation

Transgenic plants can also be produced by regeneration of hairy roots transformed with *A. rhizogenes*. Success of stable transformation has been reported in many plant species

(Christey, 2001). In soybean, stable soybean transgenic plants were produced from hairy roots using primary-node explants infected by a disarmed *A. rhizogenes* strain SHA17 (Olhoft et al., 2007) and the several reports of targeted mutation events using genome editing also have been obtained from hairy roots through *A. rhizogenes*-mediated transformation (Curtin et al., 2011; Haun et al., 2014; Demorest et al., 2016). However, genotype inflexibility has been the main hurdle for using the method in soybean.

In-Planta Gene Transformation

This is an alternative method in which *Agrobacterium* is used to infect explants, but it does not involve *in vitro* culture and regeneration of plant cells or tissues (Kalbande and Patil, 2016), thereby reducing time and labor cost, and, most importantly, avoiding somaclonal variation occurrence during *in vitro* culture-mediated genetic transformation and regeneration. In soybean, an *Agrobacterium* suspension is directly injected into the ovary (Liu et al., 2009), axillary meristematic region of germinated seedling (Chee et al., 1989), or stigma in which exogenous DNA was introduced into cells *via* the “pollen-tube-pathway” (Hu and Wang, 1999). Transgenic events could be obtained from progeny seeds. Liu et al. (2009) reported the transfer of a minimal linear marker-free and vector-free smGFP cassette into soybean by pollen tube-mediated gene transfer. Mangena (2019) summarized the progress made in *in planta* transformation and formulated a simple protocol using *in planta* *Agrobacterium* injection of seedlings. Although this could be a tissue culture bypass method and attempts for new ways are made from time to time, its efficiency has been very low and it is often not repeatable. This method has not been widely used.

Transformation Using a New Bacterium

Recently a novel bacterium, *Ochrobactrum haywardense* H1 (Oh H1), was discovered and it is capable of efficient plant transformation (Cho et al., 2022). *Ochrobactrum* is able to host for *Agrobacterium*-derived *vir* and T-DNA and helps to deliver transgenes in soybean. Oh H1-8 generated high-quality transgenic events by single-copy, plasmid backbone-free insertion at frequencies higher than those of *Agrobacterium* strains. It achieved high transformation efficiency in several soybean genotypes, which can be up to 35%. The application of the new bacterium-mediated transformation in soybean needs to be evaluated further.

PROGRESS MADE TO IMPROVE SOYBEAN TRANSFORMATION OVER THE LAST DECADE

Since 2010, increasing the transformation frequency (TF) has been the main focus for soybean transformation improvement. Several major factors affecting soybean TF based on *Agrobacterium*-mediated transformation have been identified, and progress has been made in establishing a high-throughput transformation system (Zhang et al., 2014; Arun et al., 2015, 2016; Yang X. F. et al., 2016; Li et al., 2017; Chen et al., 2018b; Karthik et al., 2020; Paredy et al., 2020). Some confirmed

TABLE 1 | Progress of soybean stable genetic transformation approaches for whole transgenic plants.

Method	Explant	Genotype	Selectable marker/agent	Physical treatment	Specific chemicals in medium	Agro-strain	Available TF (%)	References
<i>Agrobacterium</i>	Immature cotyledon	PI283332 and Peking	NptII/G418	Wounding	\	EHA101 and LBA4404	\	Parrott et al., 1989
		Jack	Hph/Hygromycin B	Wounding	AS	EHA105	0.03	Yan et al., 2000
		Jack, Williams, Ina, Macon, Dwight, and Rend	Hph/Hygromycin B	Wounding; orientation of explant (downward of the adaxial side)	AS	KYRT1	1.3 (1.1–1.7) (Jack)	Ko et al., 2003
		Mature cotyledonary node	Delmar, Maple Presto, and Peking	\	\	A208	\	Hinchee et al., 1988
			28 genotypes	Wounding and sonication	AS	KYRT1	1–2	Meurer et al., 1998
			A3237	Wounding	AS, glutamine, and asparagine	EHA101 and EHA105	0.9	Zhang et al., 1999
			12 genotypes	Wounding	\	A281, C58, ACH5, and EHA105	0.4 (one genotype)	Donaldson and Simmonds, 2000
			Bert	Wounding	AS and L-cysteine	AGL1	2.1	Olhofs and Somers, 2001
			12 genotypes	Wounding	AS, D-cysteine, and other thiol compounds	AGL1, LBA4404, GV3101, EHA105, and EHA101	\	Olhofs et al., 2001
			12 genotypes	Wounding	AS, L-cysteine, DTT, asparagine, and glutamine	EHA101	2–6.3 (glufosinate) 0–2.9 (bialaphos)	Paz et al., 2004
			Williams82	Wounding	AS and L-cysteine	EHA101	5.9	Zeng et al., 2004
			5 genotypes (Chinese soybean)	Wounding	AS, Silwet L-77, and L-cysteine, asparagine, and L-pyrogutamic acid	EHA105	3.8–11.7	Liu et al., 2008
	Half-seed	Kariyutaka	Bar/Glufosinate or Basta	Wounding (micro brush)	Silwet L-77	EHA105	4.4	Yamada et al., 2010
		PK416, JS90-41, Hara Soy, Co1, and Co2	Hph/Hygromycin B	Sonication and vacuum infiltration, wounding (hypodermic needle)	AS, DTT, L-cysteine, and sodium thiosulfate (STS)	LBA4404, EHA101, and EHA105	13.3–18.6	Arun et al., 2015
		JS-335	Bar/Glufosinate	Sonication and vacuum infiltration	AS, DTT, L-cysteine, and STS	EHA105	12.6 (10.5–16.2, J8335-bar);	Hada et al., 2018
		Jack and Zhonghuang 10	G2Epsps/Glyphosate	Sonication	Silwet L-77, AS, DTT, L-cysteine, and Na ₂ S ₂ O ₃	Ag10	2.9–5.7	Guo et al., 2015
		Jidou17	NptII/Kanamycin	Sonication	DDT, L-cysteine, sodium thiosulfate, and α -Aminoxyacetic acid	EHA105	\	Zhang et al., 2016
		7 genotypes	Bar/Glufosinate	Wounding	α -lipoic acid (α -LA), DTT, L-cysteine, AgNO ₃ , glutamine, and asparagine	EHA101	14.7	Yang J. et al., 2016
		Bert	Hph/Hygromycin B	Wounding	AS, DTT, L-cysteine, and STS	LBA4404 and EHA105	16.4 (9.4–26.2 LBA4404); 14 (9.4–26.2 EHA105)	Olhofs et al., 2003

(Continued)

TABLE 1 | (Continued)

Method	Explant	Genotype	Selectable marker/agent	Physical treatment	Specific chemicals in medium	Agro-strain	Available TF (%)	References
		Thorne, Williams, Williams79, and Williams82	Bar/Glufosinate	Wounding	AS, L-cysteine, and DTT	EHA101	3.8 (1.4–8.7)	Paz et al., 2006
		7 genotypes	Hph/Hygromycin B	Wounding (multi-needle)	AS, DTT, L-cysteine, and STS	LBA4404	\	Zhang and Xue, 2019
		5 US (Williams82) and 5 Chinese genotypes	Bar/Glufosinate	Wounding	AS, DTT, and L-cysteine	EHA105	0–6.71	Jia et al., 2015
		7 genotypes	Bar/Glufosinate	Wounding	L-cysteine and DTT	EHA105	0.5 (0–0.9)	Sato et al., 2007
		7 genotypes	Hph/Hygromycin B	Wounding (multi-needle)	AS, DTT, L-cysteine, and STS	LBA4404	\	Zhang and Xue, 2019
		5 US (Williams82) and 5 Chinese genotypes	Bar/Glufosinate	Wounding	AS, DTT, and L-cysteine	EHA105	0–6.71	Jia et al., 2015
		7 genotypes	Bar/Glufosinate	Wounding	L-cysteine and DTT	EHA105	0.5 (0–0.9)	Sato et al., 2007
		DS97–12	Hph/Hygromycin B	Sonication and vacuum infiltration	Polyamine (spermidine, spermine, and putrescine)	EHA105	29.3	Arun et al., 2016
		Williams82	Bar/Glufosinate	Wounding	AS, L-cysteine, and DDT	EHA101	1.0–3.5 (35s or NOS promoter)	Testroet et al., 2017
		8 genotypes	Bar/Glufosinate	Wounding	AS, DDT, STS, L-cysteine, AgNO ₃ , L-asparagine, L-pyroglutamic acid, and L-ascorbic acid	EHA101	7.3–10.0	Li et al., 2017
		Jack, Williams82, Zigongdongdou, and Heihe27	Bar/Glufosinate	Wounding	DTT, AS, L-asparagine, and L-glutamine	EHA101	7.6 (2.6–11.1)	Chen et al., 2018b
		DS-9712	NptII/Kanamycin	Sonication and vacuum infiltration	AS and L-cysteine	EHA105	14.51	Hada et al., 2018
		PUSA 9712	Bar/Basta	\	SNP	EHA101	34.6	Karthik et al., 2020
		Maverick and 20 proprietary elite lines	Pat/Glufosinate	Wounding	L-asparagine and L-pyroglutamic acid	EHA101 and EHA105	18.7 (12.1–23.0)	Pareddy et al., 2020
	whole cotyledonary node	ZhongHuang13	NptII/Kanamycin	Wounding	L-cysteine	EHA105	23.1	Zhang et al., 2014
	Calluses induced from either cot-node	5 genotypes	Bar/Glufosinate	\	AS, DTT, L-cysteine, and STS	AGL1	1.3 (0.3–4.3)	Hong et al., 2007
	Hypocotyls	Heinong44	NptII/Kanamycin	\	AS, L-cysteine, DTT, AgNO ₃ , and STS	EHA105	9.3	Wang and Xu, 2008
	Embryogenic cell suspension	Chapman	Hph/Hygromycin B	Sonication	AS	EHA105	\	Trick and Finer, 1998
	Embryogenic axes	P29T50, P33T50, 93Y21, DM118, and 98C21	SpcN/Spectinomycin	Sonication	AS and DDT	<i>Ochrobactrum haywardense</i> H1	35	Cho et al., 2022

(Continued)

TABLE 1 | (Continued)

Method	Explant	Genotype	Selectable marker/agent	Gene Gun/particle	Treatment	Construct form	Available TF (%) (comment)	References
Biolistic method	Immature embryo axis	Williams82 and Mandarin Ottawa	NptII/Kanamycin	Electrical, arc-discharge gun/Gold particles		Plasmid DNA		Mccabe et al., 1988
		Williams82	NptII/Kanamycin	PDS 1000/Tungsten		Plasmid DNA		Sato et al., 1993
	Somatic embryogenic suspension	Fayette	Hph/Hygromycin B	DuPont Biolistics TM Particle Delivery System (Model BPG)/Tungsten particles		Plasmid DNA	0.4	Finer and Mcmullen, 1991
		Fayette	Npt II/G418	PDS 1000/Tungsten particles		Plasmid DNA	Four plants per bombarded flask	Sato et al., 1993
		Fayette	Hph/Hygromycin B	PDS 1000/Tungsten particles		Multiple plasmid DNA	\(co-transformation)	Hadi et al., 1996
		\	Hph/Hygromycin B	PDS 1000/Tungsten particles		Plasmid DNA	\(protocol)	Finer and Larkin, 2008
		\	Hph/Hygromycin B	PDS 1000/Tungsten particles		Plasmid DNA	\(protocol)	Finer, 2016
		93B86	Hph/Hygromycin B and Als/Chlorsulfuron	PDS 1000/Gold particles		Plasmid DNA and DNA fragment	\(targeted insertion)	Li et al., 2015
	Mature embryo axis	BR-16, Doko RC, BR-91, and Conquista	AHAS/Imazapyr	HPHMAS/Tungsten	\	Plasmid DNA	0.1–7.8	Aragão et al., 2000
		BR-16, BR-91, Celeste, Conquista, Doko RC, Nina, Indiana, and Itaipu	AHAS/Imazapyr	PDS1000/Tungsten		Plasmid DNA	≤0.2 (protocol)	Rech et al., 2008
		Conquista	AHAS/Imazapyr	HPHMAS/Tungsten	\	DNA fragments	0.8	Vianna et al., 2011
	Immature embryo	INCASoy-36	Cp4epsps/Glyphosate	PDS 1000/Tungsten		Plasmid DNA	6	Soto et al., 2017
		Maverick	Hph/Hygromycin B, DSM2/Glufosinate		Cold treatment and plasmolysis	Plasmid DNA	2–5.5 (hph) and 1–2.7 (DSM2)	Chennareddy et al., 2018

(1) HPHMAS: The high-pressure helium-driven microparticle acceleration system. (2) \ means not available. (3) Protocol means the reference is a published protocol.

positive elements in *Agrobacterium*-mediated transformation protocols have also been applied for enhancing soybean biolistic transformation (Table 1).

Agrobacterium-Mediated Transformation

Soybean transgenic plant production still relies on *Agrobacterium*-mediated transformation (Figure 2 and Table 1). Recently, high TFs of over 10% have been obtained in more and more soybean genotypes using improved protocols (Zhang et al., 2014; Arun et al., 2015, 2016; Yang X. F. et al., 2016; Li et al., 2017; Chen et al., 2018b; Karthik et al., 2020; Pareddy et al., 2020). The enhancement of TF is based on changes in several factors, including explant, selectable marker, and culture medium composition such as antioxidants, of these protocols (Table 1).

Adjustment of Infection Method and Improving Regeneration

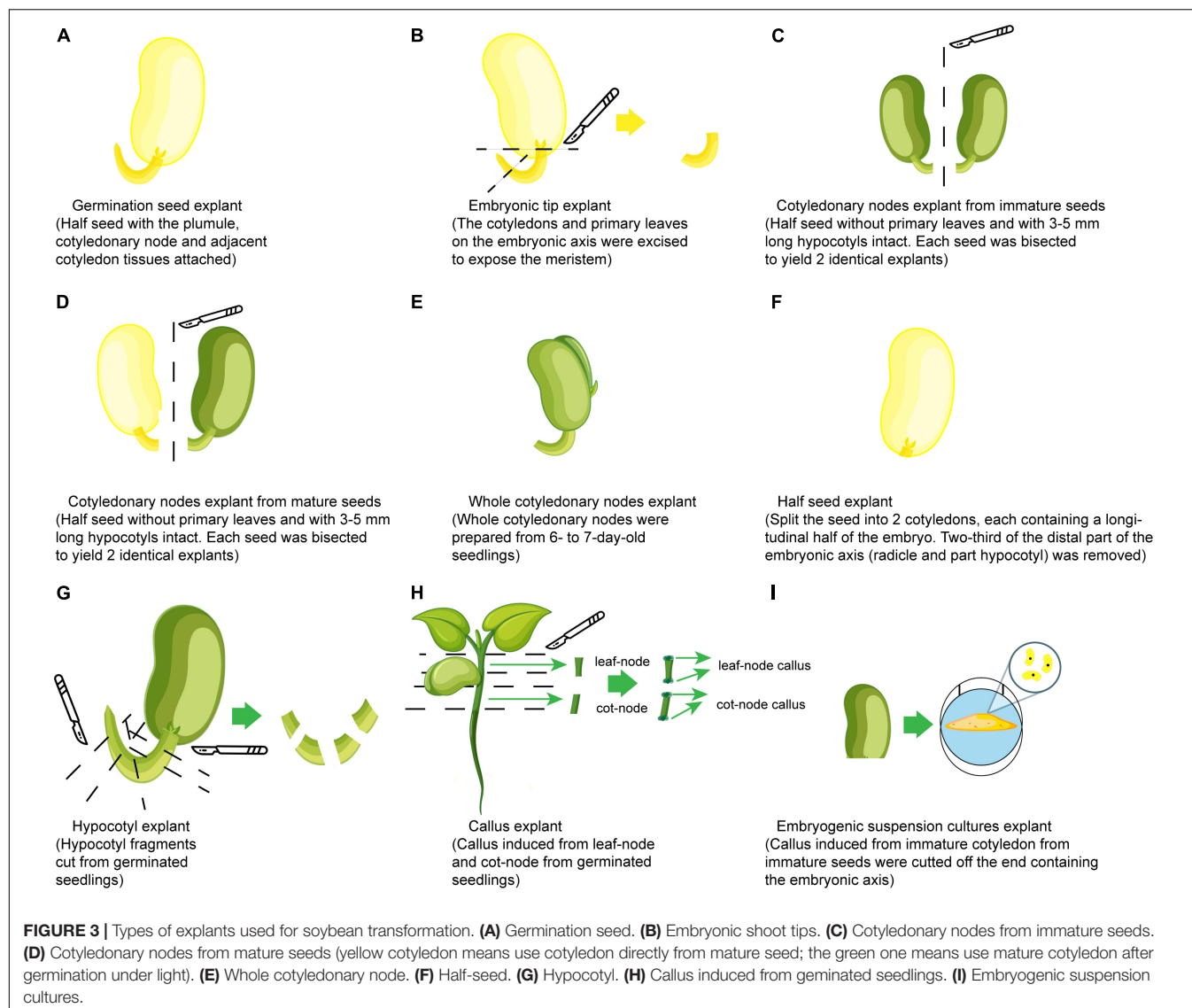
Reducing the explant tissue browning and necrosis caused by *Agrobacterium* enhances construct delivery and regeneration of transformed cells. Changing the ways for preparation of *Agrobacterium* infection solutions and co-cultivation media, and modifying infection methods can achieve this goal and eventually increase transformation efficiency. Addition of antioxidants such as dithiothreitol (DTT) in infection solutions and extending co-cultivation time to 5 days achieved an infection efficiency of more than 96% and, hence, increased TF (Li et al., 2017). Infection solutions prepared with a two-round overnight culture of *Agrobacterium* using AB minimal media in second round culture significantly increased transformation frequency in comparison with the culture using normal YEP medium (Pareddy et al., 2020). It was also found to be beneficial to *A. tumefaciens* infection when the co-cultivation temperature for soybean transformation was set to 23°C under dim light (Yang X. F. et al., 2016). The same group also demonstrated to alleviate explant necrosis and significantly improve the transformation efficiency when antioxidants alone such as α -lipoic acid (α -LA, 0.12 mM) and silver nitrate (AgNO₃, 20 μ M), or combinations of antioxidants such as L-cysteine (1 mM) + DTT (3.3 mM) + AgNO₃ (20 μ M), and L-cysteine (1 mM) + DTT (3.3 mM), were added in the solid co-cultivation medium. For improving regeneration, it was found that adding 6-benzylaminopurine (BAP) in a germinating medium could significantly increase regeneration efficiency, which led to enhancement of TF; the optimal BAP concentration for shoot formation was 0.5 mg/L (Zhang et al., 2014). More examples are presented in Table 1.

Genotype Effect and Explant Choice

In the tissue culture-based transformation process, the composition of culture media and susceptibility of selected explants to *Agrobacterium* influence soybean transgenic frequency. A highly efficient *in vitro* culturing system and regeneration of cells susceptible to *Agrobacterium* are prerequisites for a reliable transformation protocol. Until now, the TF for most tested genotypes of soybean has remained quite low at a level mostly below 5% when conducting *Agrobacterium*-mediated transformation [summarized in Yamada et al. (2012), Jia et al. (2015), and Li et al. (2017); Table 1]. Since 2000, many

research groups have used model soybean varieties such as Jack, Bert, and Williams serials and other specific genotypes because of their amenability to transformation (Olhoft and Somers, 2001; Olhoft et al., 2001, 2003; Paz et al., 2004, 2006; Zeng et al., 2004; Luth et al., 2015). Recently, soybean transformations with high TFs have been reported using specific genotypes. For example, it was claimed 23.1% with Zhonghuang13 (Zhang et al., 2014) and an average of 14% TF for a local Indian genotype, DS-9712 (Hada et al., 2018). Improvement based on *Agrobacterium*-mediated soybean transformation has been made to expand target genotypes from conventional model varieties to many elite varieties (Ko et al., 2003; Yi and Yu, 2006; Sato et al., 2007; Song et al., 2013; Arun et al., 2015; Pareddy et al., 2020). For example, over 5% TF for more than 10 varieties was achieved with a robust protocol (Pareddy et al., 2020).

Since Hinchee et al. (1988) obtained transgenic events, the cotyledonary node of mature seeds has been the most favorite explant used for *Agrobacterium*-mediated soybean transformation using many other explants such as embryonic tips and calli (Figure 3). Cotyledonary node regions have axillary meristems at the junction between cotyledon and hypocotyl, which can proliferate and regenerate by the formation of multiple adventitious shoots on a culture medium containing cytokinin. Successful transformation has been achieved using similar organogenesis from various explants, which include germination seeds (Chee et al., 1989), embryonic shoot tips (Martinell et al., 2002; Liu et al., 2004), cotyledonary nodes from immature seeds (Parrott et al., 1989; Yan et al., 2000; Ko et al., 2003), cotyledonary nodes from mature seeds (Meurer et al., 1998; Zhang et al., 1999; Donaldson and Simmonds, 2000; Olhoft and Somers, 2001; Olhoft et al., 2001; Paz et al., 2004; Zeng et al., 2004; Liu et al., 2008), half-seeds (Paz et al., 2006; Pareddy et al., 2020), whole cotyledonary nodes (Zhang et al., 2014) and hypocotyls (Dan and Reichert, 1998; Liu et al., 2004; Wang and Xu, 2008), and other explants with different regeneration procedures such as calli induced from germinated seedlings (Hong et al., 2007) and embryogenic suspension cultures (Trick and Finer, 1998). However, successful and repeatable production of transgenic soybean via *Agrobacterium*-mediated transformation has mainly been based on protocols with explants containing cotyledonary nodes from young seedlings and imbibed mature seeds (Zhang et al., 1999; Olhoft et al., 2003; Paz et al., 2006). Recently, half-seeds have gradually become the trend for explants since (Paz et al., 2006) their first use, because half-seed explants possess advantages to have more nutrition supply for shoot regeneration compared to cotyledonary nodes and to be prepared within a short time (less than 1 day) due to using imbibed seeds, which reduces the period of total regeneration and labor cost. Based on descriptions of explants in several reports (Paz et al., 2006; Pareddy et al., 2020), half-seed, whole cotyledon, and split seed explants can now be put under the same category of half-seed explants. Obtaining TFs of over 10% for soybean with half-seed explants have been demonstrated in many reports (Zhang et al., 2014; Arun et al., 2016; Li et al., 2017; Chen et al., 2018b; Hada et al., 2018) (Table 1). The highest TF of 34.6% has been obtained using these explants together with nitric oxide treatment in a co-cultivation medium in the protocol made by



Karthik et al. (2020). Some specific explant treatments such as sonication in combination with vacuum infiltration, sonication in combination with surfactant, or just sonication (Mariashibu et al., 2013; Arun et al., 2015; Guo et al., 2015; Zhang et al., 2016; Hada et al., 2018), and pre-wounding with a multi-needle consisting of 30 thin fibers (Xue et al., 2006) or a micro-brush (Yamada et al., 2010) were also used before *Agrobacterium* infection to increase infection rate and TFs, because these treatments facilitate the penetration of *Agrobacterium* into plant tissues and increase the contact between plant cells and the bacterium, and stimulate the infection ability of the bacterium, which leads to T-DNA transfer into plant cells.

Addition of Antioxidants in Medium

Antioxidants, in general, are known to reduce pathogen-induced programmed cell death (Mittler et al., 1999). These include inhibitors of polyphenol oxidases (PPOs) and peroxidases (PODs) through the action of their thiol group, such as

compounds L-cysteine, DTT, and sodium thiosulfate. They are commonly used to reduce enzymatic browning in food processing caused by deposition of tannins (Nicolas et al., 1994; Ghidelli et al., 2014). Polyvinylpyrrolidone (PVP), DTT, L-cysteine, glutathione, α -LA, L-ascorbic acid, and citric acid have been confirmed to decrease tissue necrosis of explants used for *Agrobacterium*-mediated transformation (Barampuram and Zhang, 2011). Either one or more than 2 of the chemicals have been used in soybean transformation (Olhoft and Somers, 2001; Olhoft et al., 2003; Paz et al., 2004; Yi and Yu, 2006; Liu et al., 2008). L-cysteine and DTT have been frequently used in soybean transformation since its first use by Olhoft et al. (2001). Reports clearly showed that there was less browning on the cut and damaged surfaces of the hypocotyl, cotyledon node region, and on the cotyledon of explants, which increased the TF of stable transformations (Olhoft et al., 2001, 2003; Paz et al., 2004; Li et al., 2017). A high average TF of 12.7% resulted from the combination of L-cysteine and

DTT, which was significantly greater than that of either L-cysteine or DTT alone (Olhoft et al., 2003). The positive effect has been continuously confirmed in recent reports (Table 1). Another type of antioxidants is a group of sulfur-containing compounds involved in several multienzyme complexes such as α -LA. These include pyruvate dehydrogenase, α -ketoglutarate dehydrogenase, branched-chain ketoacid dehydrogenase, and glycine decarboxylase (Dan et al., 2009). Adding the antioxidant α -LA in a co-cultivation medium could increase transient GUS expression and increased the percentage of shoot induction (Yang X. F. et al., 2016). In this report, 0.12 mM α -LA was found to be the most useful for alleviating browning and necrosis. Other antioxidants conventionally used in plant tissue culture, such as ascorbic acid, PVP, and citric acid, may promote soybean transformation efficiency, but their roles have not yet been made clear (Li et al., 2017). Plant hormone-like antioxidants such as sodium nitroprusside (SNP), a nitric oxide (NO) donor, play varied roles in growth and development of plants. Nitric oxide is involved in cell metabolism and morphogenesis and acts as a signaling molecule in response to various biotic and abiotic stresses (Verma et al., 2020), and can alleviate abiotic stress threat in plants reacting quickly with ROS. SNP significantly enhanced regeneration and development rate of soybean plants (Karthik et al., 2019); addition of SNP also significantly increased soybean TF by up to 34.6% with the half-seed method (Karthik et al., 2020).

Addition of Other Chemicals in Culture Medium

Except for antibiotics, chemicals related to host defense response, ethylene inhibitors, surfactants, demethylating reagents, polyamines, and antagonist α -aminooxyacetic acid (AOA) are proved to have a positive effect on improving TF. L-glutamine and L-asparagine are types of chemicals that weaken host defense responses. It has been reported that the addition of L-glutamine into a culture medium alone or in combination with a cold shock pretreatment could enhance *Agrobacterium* transformation efficiency (Zhang et al., 2013). Although the mechanism is still not clear, L-glutamine could play a role in lessening host defense responses by attenuating the expression of certain pathogenesis-related genes (PRs), and potentially improve the efficiency of *Agrobacterium*-mediated plant transformation (Zhang et al., 2013, 2014). It was demonstrated that TF was significantly increased in soybean when additional L-glutamine or L-asparagine alone, or both of them were added in all culture media (Chen et al., 2018b). The TF was 8.8 ± 1.5 (L-glutamine), $5.9 \pm 2.1\%$ (L-asparagine), 11 ± 0 (both), and $3.5 \pm 2.4\%$ (without any one of them). Ethylene inhibitors such as AgNO₃ have a positive effect on transformation. It has been reported that Ag⁺ interferes with the binding of ethylene receptor sites and helps reduce ethylene production by promotion of polyamine biosynthesis (Roustan et al., 1990). The main function of AgNO₃ is to eliminate the potential danger to plant cells and tissues in liverwort caused by ethylene (Beyer, 1979). It has already been confirmed to promote somatic embryo production and shoot regeneration in wheat and maize (Carvalho et al., 1997; Fernandez et al., 1999). This effect has been proved to improve soybean TF (Olhoft et al., 2004; Li et al., 2017). A nearly 10%

TF with genotype Heilong44 was reported when BAP and AgNO₃ were added into a culture medium (Wang and Xu, 2008). Surfactants such as SilwetL-77 and pluronic acid F68 also increase TF, which initially showed to enhance T-DNA delivery in wheat *Agrobacterium*-mediated transformation when added into an inoculation medium (Cheng et al., 1997). This was also confirmed in soybean transformation. It was reported that adding SilwetL-77 to an infection medium coupled with hygromycin-based selection strategies led to transformation efficiencies ranging from 3.8 to 11.7% in Chinese soybean varieties (Liu et al., 2008). SilwetL-77 has been frequently used to increase soybean TF (Yamada et al., 2010; Guo et al., 2015). Surfactants may enhance T-DNA delivery by aiding *A. tumefaciens* attachment and/or by elimination of certain substances that inhibit *A. tumefaciens* attachment (Opabode, 2006). Polyamines enhance plant cell differentiation, induce totipotency, and increase cell division (Rakesh et al., 2021). Addition of polyamines in the plant transformation process leads to *vir* gene induction and T-DNA transfer, and increases transformation efficiency (Kumar and Rajam, 2005). As high as 29.3% TF in soybean has been achieved by addition of spermidine, spermine, and putrescine in a culture medium compared with its counterparts (14.6%) and with respective plant growth regulator (PGR) alone (Arun et al., 2016). Demethylating reagents commonly applied in epigenetic research such as 5-azacytidine (5-Azac), significantly improve the transient transfection efficiency and transgene expression level in low-efficiency genotypes. Treatment with 5-Azac improved the shoot regeneration efficiency in low-efficiency genotypes during the process of *Agrobacterium*-mediated soybean transformation. This indicates that lower methylation level in transgenes contributed to enhance shoot regeneration in *Agrobacterium*-mediated soybean transformation (Zhao et al., 2019b). Antagonist AOA relieves the structural membrane barriers of *Agrobacterium* entering cells, hinders the perception of intercellular signal transmission, and thus effectively alleviates defense responses and increases the susceptibility of cells to *Agrobacterium* infection. Combined use of AOA and sonication treatments (novel method) greatly improved T-DNA delivery efficiency in soybean (Zhang et al., 2015, 2016).

Refining Selection Agents

The most frequently used selectable markers in both the somatic embryogenesis- and organogenesis-based soybean transformation methods are genes conferring resistance to herbicides or antibiotics so as to reduce escape rate significantly. The selectable markers include *bar* and *pat* genes conferring resistance to phosphinothricin, the active ingredient in BASTA and bialaphos herbicides (Zhang et al., 1999; Olhoft and Somers, 2001; Olhoft et al., 2001; Paz et al., 2004; Testroet et al., 2017; Pareddy et al., 2020), EPSPS (5-enolpyruvylshikimate-3-phosphate synthase) genes conferring resistance to the herbicide glyphosate (Martinell et al., 1999; Clemente et al., 2000; Yao, 2001; Guo et al., 2015; Xiao et al., 2019), and the *nptII* gene conferring resistance to the antibiotics kanamycin (Homrich et al., 2012) and *hph* or *hpt II*

(hygromycin phosphotransferase) genes conferring resistance to hygromycin B (Yan et al., 2000; Ko et al., 2003; Olhoft et al., 2003; Liu et al., 2008). Recently, *Hph* and *Bar* or *Pat* have been proven to be the most favorite selectable markers (Table 1). An average transformation frequency as high as 29.3% was achieved with the half-seed (Arun et al., 2016) and 13.3–18.6% with cotyledonary node explant (Arun et al., 2015) employing hygromycin B selection, which is better than or comparable with that by Olhoft et al. (2003). Glufosinate has also been used as a selection agent based on the *bar* or the *pat* gene and initially had less than 10% TF in soybean transformations involving the half-seed (Paz et al., 2006) and embryo tip (Dang and Wei, 2007) explants. Recently TFs of over 10% have been obtained with cotyledonary nodes (Hada et al., 2014; Yang X. F. et al., 2016) and half-seeds (Li et al., 2017; Chen et al., 2018b; Pareddy et al., 2020). A 34.6% TF was reported using a protocol with addition of sodium nitroprusside (SNP) when Basta was sprayed for selection (Karthik et al., 2020). Since an *epsps*/glyphosate selection based protocol is established (Martinell et al., 1999), glyphosate has gradually been incorporated into transformation as a selectable agent and has shown its beneficial side for high stringency. In order to quickly and efficiently screen glyphosate-tolerant events, a rapid and convenient spotting method was established for screening regenerated glyphosate-tolerant T₀ plantlets (Guo et al., 2020a). In this report, an optimized *Agrobacterium*-mediated soybean transformation system with rapid and effective selection of transformed cells was developed, with TFs ranging from 2.9 to 5.6%. Especially, 96% regenerated T₀ plantlets showed clear tolerance to glyphosate and their transgenic nature were confirmed by molecular analysis. Spectinomycin was also used as a selective agent to obtain transgenic soybean when the aminoglycoside-3'-adenylyltransferase gene (*aadA*) was used as a selectable marker (Martinell et al., 2002). The spectinomycin selection protocol demonstrated higher frequency of transformation, a shorter period of time needed to complete each protocol, and lower cost compared with the glyphosate selective protocol. Soybean transformation using a GFP as a detection marker was also reported, and transgenic plants could be identified at an early stage, although the frequency was not high (2.5%) (Yang S. et al., 2019). Combined with a normal selectable marker and a selection agent, GmFAST (fluorescence-accumulating seed technology) has recently been developed to identify homologous transgenic seeds. It is a marker composed of a soybean seed-specific promoter coupled to the OLE1-GFP gene, which encodes a GFP fusion of the oil-body membrane protein OLEOSIN1 of *Arabidopsis thaliana* and is a time-saving and efficient method to produce homologous transgenic events (Iwabuchi et al., 2020).

Generally, the efficiency of *Agrobacterium*-mediated transformation in soybean has been enhanced by improving both *Agrobacterium* infection and explant regeneration. Addition of antioxidants such as DDT, L-cysteine, and NO in a co-cultivation medium and some infection-assisting specific chemicals including surfactants and AgNO₃, and some regeneration-promoting elements such as polyamines (Table 1), plays an

important role in the improvement. All of the measures have facilitated *Agrobacterium* to transform the meristematic region of soybean explants. Recently, a specific *Agrobacterium*-mediated protocol was reported, which conducted bombardment to make wounding and reduce the lab work time to only 2 days in the transformation process and kept the rest time to grow T₀ plants in a glasshouse (Paes de Melo et al., 2020). Transgenic events were screened using a swab to spread glufosinate solution on leaves of putative events and the TF reached nearly 10%. This method avoided many tissue culture steps and may be a cost-saving protocol.

Biolistic Transformation

Since McCabe et al. (1988) reported the first transgenic soybean plant using the biolistic method, many reports of soybean biolistic transformation have been published and the development of this method in the first 25 years has been reviewed by Homrich et al. (2012), Lee et al. (2013), Mariashibu et al. (2013), and Mangena et al. (2017). Initially, meristems of soybean tissues as the target tissue were used for bombardment such as embryonic axes of immature and mature seeds (Sato et al., 1993; Aragão et al., 2000; Rech et al., 2008; Soto et al., 2017). In later studies, somatic embryos (Finer and McMullen, 1991; Finer and Larkin, 2008; Finer, 2016) were the most frequently used explants for biolistic transformation. However, chimeric transgenic plants were produced because of multiple cell layers (L1, L2, and L3) in the original apical meristem of soybean (Christou, 1990; Christou and McCabe, 1992). Fortunately, using secondary somatic embryos and new selective markers such as EPSPS has eliminated transgenic chimeras (Sato et al., 1993; Martinell et al., 1999). Somatic embryo regeneration and proliferation were initiated either on semi-solid media (Parrott et al., 1989) or liquid suspension culture media (Finer and Nagasawa, 1988). Co-transformation of multiple plasmids or multiple gene inserts in same constructs with selectable markers has been achieved (Hadi et al., 1996; Li et al., 2015). Since 2010, factors such as explant type, abiotic stress treatment, selectable marker, and tissue culture method have been the main focus to improve biolistic transformation TF, and reliable protocols for the biolistic method with embryogenesis-based explant have been developed to produce a reasonable number of transgenic plants (Table 1). For example, a TF of up to 6% was achieved with *cp4epsps* as selectable marker when embryonic axes of mature seeds of the INCASoy-36 Cuban cultivar were bombarded (Soto et al., 2017). Chennareddy et al. (2018) combined an immature half-seed explant with an intact embryonic axis, cold and plasmolysis pre-treatment, and a specific somatic embryogenic callus regeneration medium in their protocol. They achieved 5% TF with HPH/hygromycin selection and 2.7% with DSM2/glufosinate selection. A selection system using NPTII/G418 was developed for a biolistic-transformed embryogenic callus rather than the most used HPH/hygromycin system and similar TF in comparison with the HPH system was obtained (Itaya et al., 2018). The current status is that soybean biolistic transformation still relies on an embryogenic callus, since it is the prerequisite for establishing a robust

transformation system for a specific genotype. Selection for the amenability of an embryogenic callus induced from local elite varieties (genotypes) is the main focus (Joyner et al., 2010; Abbasi et al., 2016; Islam et al., 2017; Raza et al., 2020). An improved biolistic soybean transformation protocol was published using an embryogenic callus induced from an immature cotyledon explant (Finer, 2016), which is a robust one and can produce quite a lot of transgenic plants within 6–9 months.

RECENT APPLICATIONS OF SOYBEAN TRANSFORMATION FOR TRAIT IMPROVEMENT

Transgenic technology has been used to improve soybean agronomic traits, which include yield component, grain quality, and biotic and abiotic stress tolerance, and economic traits such as oil and biofuel quality, and specific chemical content in seed for human health, and other traits. Trait improvements through forward and reverse genetic approaches in the last 5 years are summarized in **Table 2**; i.e., downregulation of the pyruvate dehydrogenase kinase gene *GmPDHK* through RNAi made an average of 42.2% protein content in seeds of transgenic plants, which is significantly increased compared with the non-transgenic control (Jones et al., 2020). Soybean seeds with linolenic acid content in excess of 50% of the total oil have been generated by increasing the expression of the *FAD3* gene, which encodes the enzyme that converts linoleic acid to linolenic acid (Yeom et al., 2020). Overexpressing the *GmmiR156b* (Squamosa promoter-binding protein-like, SPL) gene in soybean and transgenic plants produced significantly increased numbers of long branches, nodes, and pods that exhibited increased 100-seed weight, resulting in a 46–63% increase in yield per plant and no significant impact on plant height in a growth room or under field conditions (Sun et al., 2019). Stable *GmMYB14*-overexpressing (*GmMYB14*-OE) transgenic soybean plants demonstrate semi-dwarfism and a compact plant architecture associated with decreased cell size, causing decreased plant height, internode length, leaf area, leaf petiole length, and leaf petiole angle, and improved yield in high density and drought tolerance under field conditions (Chen et al., 2021b). Salt-tolerant transgenic soybean and its applications in field are summarized in a review (Cao et al., 2018). Resistance to soybean cyst nematode (SCN; *Heterodera glycines*) in stably transformed soybean plants is enhanced by downregulation of the *HgY25* and *HgPrp17* genes, which are related to reproduction and fitness (Tian et al., 2019). Overexpression of *PAC1* and *GmKR3*, a TIR–NBS–LRR-type R gene, can increase multiple virus resistance in transgenic soybean and, thus, provide an efficient control strategy against RNA viruses such as SMV, BCMV, WMV, and BPMV (Xun et al., 2019). Overexpression of *GmDRI1* [*Glycine max* disease resistance 1 (*Glyma.10G094800*)] led to enhanced resistance not only against *F. virguliforme* but also against spider mites (*Tetranychus urticae*, Koch), soybean aphids (*Aphis glycines*, Matsumura), and SCN (Ngaki et al., 2021). Many types of

herbicide-resistant transgenic soybean, such as glyphosate-resistant, dicamba-, and 2,4-D-resistant, are grown widely in the United States (Nandula, 2019). Transgenic soybean plays an important role in soybean production worldwide now, and transgenic soybean covers 50% of the global transgenic crop area, occupying 94.1 million ha (Nandula, 2019). Therefore, a better soybean transformation system is the base for soybean improvement through transgenic technology.

CHALLENGES AND FUTURE DIRECTIONS IN SOYBEAN TRANSFORMATION

Although much effort has been made to improve the transformation systems for soybean, there are some challenges such as genotype flexibility, low transformation frequency, time to time chimerism in T₀ transgenic plants, and availability of a system for new breeding technologies such as genome editing.

Genotype Flexibility

Like in other recalcitrant plant species, genotype inflexibility has been an obstacle that restricted the scope of soybean transformation. The ideal soybean transformation target material for trait improvement would be any elite variety with excellent agronomic characteristics. However, most reliable transformations are still based on specific genotypes although genotypes amenable to transformation have expanded to some preferred genotypes. For example, in the early stage, successful *Agrobacterium*-mediated transformation occurred in several genotypes and their derivatives such as Williams, Williams79, and Williams82 (Paz et al., 2004, 2006). High-efficiency *Agrobacterium*-mediated transformation is only achieved in a limited number of elite lines (Zhang et al., 2014; Arun et al., 2015, 2016; Yang J. et al., 2016; Li et al., 2017; Chen et al., 2018b). High-efficiency genotypes possess greater susceptibility to *Agrobacterium* infection, which has been confirmed in many reports (Jia et al., 2015; Yang J. et al., 2016; Yang X. F. et al., 2016; Zhao et al., 2019b). The competency of cotyledons of seeds to *Agrobacterium* infection and the ability to regenerate plants are key factors. These may be determined by cell defense response, including attachment of *A. tumefaciens* to plant cells, plant signals sensed by *A. tumefaciens*, regulating *vir* gene expression, T-DNA/virulence protein transport or initial contact of *A. tumefaciens* to plants and cytoplasmic trafficking, and nuclear import of T-DNA and effector proteins (Hwang et al., 2017). An important step to enhance the transformation efficiency of recalcitrant genotypes is to improve the genotypes' susceptibility to *Agrobacterium* infection. Many commonly used treatments to increase transformation efficiency such as heat shock, cold shock, antioxidants, and hypoxia may act by suppression of cellular response to *Agrobacterium* infection (Zhang et al., 2013). Combinations of various positive factors discovered or developed recently have promoted *Agrobacterium*-mediated soybean transformation to extend genotype scope (**Table 1**). For example, transgenic events have been obtained from 19 out of 20 genotypes based on an improved protocol

TABLE 2 | Summary of transgenic approaches for soybean trait improvement and functional genomics in the last 5 years.

Target traits	Transgene	Source of gene	Delivery method	Effect on trait or function	Genotype	References
Seed components and quality						
Seed protein and amino acid	<i>Zmδ-zeins</i> and <i>Zmγ-zein</i>	<i>Z. mays</i>	<i>A. tumefaciens</i>	Increase 27% the methionine content	Williams82	Kim and Krishnan, 2019
	<i>Zmβ-zein</i>	<i>Z. mays</i>	<i>A. tumefaciens</i>	Increase 15% the methionine content	Jack	Guo et al., 2020b
Oil	<i>GmPDHK</i>	<i>G. max</i>	Biolistic method	Increase average 42.2% protein content	Jack	Jones et al., 2020
	<i>Glyma.10G38760a</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase sulfur amino acid content	Maverick	Kim et al., 2020
	<i>GmFAD2-1B</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase oleic acid content	Williams82	Yang J. et al., 2018
	<i>GmSDP1-1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase oil content	Kariyutak	Kanai et al., 2019
	<i>PtFAD3-1</i>	<i>P. fendleri</i>	<i>A. tumefaciens</i>	Increase α-linolenic acid production	Kwangankong	Yeom et al., 2020
	<i>GmOLEO1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase 10.6% seed oil content and enriched smaller OBs	Williams82	Zhang et al., 2019a
	<i>GmWRI1b</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increases total seed oil production	\	Guo et al., 2020c
	<i>Glyma.13G30950</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase seed pods and oil production	Kariyutaka	Iwabuchi et al., 2020
	<i>GmDGAT2A</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase oil production and α-linoleic acid content	P03	Jing et al., 2021
	<i>GmZF392</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase seed oil accumulation	Jack	Lu et al., 2021
	<i>GmWRI1a</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase seed oil content	Dongnong50	Wang Z. et al., 2022
	<i>AhDGAT3</i>	<i>A. hypogaea</i>	<i>A. tumefaciens</i>	Increase oleic acid and total fatty acid	Jack	Xu et al., 2022
Bioreactor	<i>rhBMP2</i>	<i>H. sapiens</i>	Biolistic method	Result in production of bone morphogenetic protein BMP2	BRS16	Queiroz et al., 2019
	The lunasin gene	<i>G. max</i>	<i>A. tumefaciens</i>	Result in production of bioactive lunasin peptide	\	Hao et al., 2020
	The <i>hIFN-γ</i> gene	<i>H. sapiens</i>	<i>A. tumefaciens</i>	Result in production of human IFN-γ protein	Williams	Mehrizadeh et al., 2021
Phytate content	<i>GmIPK2</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in production of low phytate	Pusa-16	Punjabi et al., 2018
	<i>GmMIPS1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Regulate phytate biosynthesis	DS-9712	Kumar et al., 2019
	<i>EcMappA</i>	<i>E. coli</i>	<i>A. tumefaciens</i>	Result in production of a thermostable phytase	Wandou-28	Zhao et al., 2019c
Specific chemical compounds	<i>ZmGB1</i>	<i>Z. mays</i>	<i>A. tumefaciens</i>	Increase glycinebetaine content	A5403, A4922, A3469, and A3244	Castiglioni et al., 2018
	<i>GmCHI1A</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase seed isoflavones	DT2008	Nguyen et al., 2020
	<i>GmMATE1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase seed isoflavones	C08 and W05	Ng et al., 2021
	<i>GmMYB176</i> and <i>GmbZIP5</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase seed isoflavones	Harosoy63	Anguraj Vadivel et al., 2021
Agronomic traits						
Seed yield and plant biomass	<i>GmPT7</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase symbiotic N ₂ fixation and yield	HN66	Chen et al., 2019b
	<i>GmmiR156b</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Improve the shoot architecture and yield	Williams82	Sun et al., 2019
	<i>psNTP9</i>	<i>P. sativum</i>	<i>A. tumefaciens</i>	Increase soybean yield	Williams82	Veerappa et al., 2019
	<i>GmWRI1b</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Improve plant architecture and associated yield parameters, and increases total seed oil production	\	Guo et al., 2020c
	<i>HaHB4</i>	<i>H. annuus</i>	<i>A. tumefaciens</i>	Enhance drought tolerance with yield reduced	Williams82	Ribichich et al., 2020
	<i>GmMYB14</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance high-density yield and drought tolerance	Tianlong1	Chen et al., 2021b
	<i>ZmSOC1</i>	<i>Z. mays</i>	<i>A. tumefaciens</i>	Increase soybean yield	Jack	Han et al., 2021
	<i>GmFULa</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase soybean yield	Zigongdongdou	Yue et al., 2021
	<i>GmHSP17.9</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase nodule number, nodule fresh weight, and seed yield	Williams82	Yang et al., 2022

(Continued)

TABLE 2 | (Continued)

Target traits	Transgene	Source of gene	Delivery method	Effect on trait or function	Genotype	References
Plant architecture	<i>GmIDL2a and GmIDL4a</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase the lateral roots densities of the primary roots	XIAOLIDOU	Liu C. et al., 2018
	<i>GmmiR156b</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Improve the shoot architecture and yield	Williams82	Sun et al., 2019
	<i>GmYUC2a</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Delay nodule development and a reduced number of nodules	Williams82	Wang et al., 2019c
	<i>GmGASA32</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase plant height	Williams82	Chen et al., 2020b
	<i>GmWRI1b</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Improve plant architecture and associated yield parameters, and increases total seed oil production	\	Guo et al., 2020c
	<i>Glyma.13G30950</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase seed pods and oil production	Kariyutaka	Iwabuchi et al., 2020
	<i>GmPIF4b</i>	<i>G. max</i> var. Bragg	<i>A. tumefaciens</i>	Affect plant morphology and accelerating reproductive phase transitions	Bragg	Arya et al., 2021
	<i>AtBIC1</i>	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Increase plant height	Kwangankong	Cho et al., 2021
	<i>GmDIR27</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase pod dehiscence	Williams82	Ma X. et al., 2021
	<i>GA2ox8A and GA2ox8B</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Decrease trailing growth and shoot length	W05	Wang et al., 2021d
	<i>GmGAMYB</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Promote flowering and increase plant height	DongNong50	Yang et al., 2021
	<i>GmBICs</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase stem elongation	TianLong1	Mu et al., 2022
	<i>GmbHLH57 and GmbHLH300</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance Fe uptake and increase the Fe content in plants	Williams82	Li et al., 2018
	<i>GmPT7</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance symbiotic N ₂ fixation and yield	HN66	Chen et al., 2019b
	<i>GmWRI1s</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase nodule numbers	Tianlong1	Chen et al., 2020a
Iron, nitrogen, and phosphorus use efficiency	<i>GmPAP12</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase nodule numbers	Williams82	Wang et al., 2020d
	<i>GmAAP6a</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance tolerance to low nitrogen and improve seed nitrogen status	Tianlong1	Liu et al., 2020
	<i>GmMDH12</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Decrease nodule size and mediates malate synthesis	YC03-3	Zhu et al., 2021
	<i>GmNMHC5</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase nodulation	Jack	Wang W. et al., 2022
	<i>GmNINs</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Decrease nodule numbers	Williams82 and Huachun6	Fu et al., 2022
	<i>GmD27c</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase nodule numbers	Tianlong1	Rehman et al., 2022
	<i>GmSPX8</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase nodule number, nodule fresh weight, and nitrogenase activity	Zhonghuang15	Xing et al., 2022
	<i>GmHSP17.9</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase nodule number, nodule fresh weight, and seed yield	Williams82	Yang et al., 2022
	<i>EsPHT1;4</i>	<i>E. salsugineum</i>	<i>A. tumefaciens</i>	Increase tolerance to low phosphorus stress	YD22	Yang et al., 2020b
	<i>GmETO1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance Pi deficiency tolerance	NN94156 and Bogao	Zhang H. et al., 2020
	<i>GmFT1a and GmFT2a/5a</i>	<i>G. max</i>	<i>A. tumefaciens</i>	<i>GmFT1a</i> and <i>GmFT2a/5a</i> have opposite roles in controlling flowering	Zigongdongdou and Heihe27	Liu W. et al., 2018
	<i>GmFT2b</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Promote flowering	Jack	Chen et al., 2020c
	<i>GmGAMYB</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Promote flowering and increase of plant height	DongNong50	Yang et al., 2021
	<i>E1 (Glyma06g23026)</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Promote flowering	Zigongdongdou	Liu et al., 2022a
Abiotic and biotic traits						
Nematode resistance	<i>HgY25</i>	<i>H. glycines</i>	Biostistic method	Enhance resistance to soybean cyst nematodes	Jack	Tian et al., 2019

(Continued)

TABLE 2 | (Continued)

Target traits	Transgene	Source of gene	Delivery method	Effect on trait or function	Genotype	References
Insect resistance	<i>BtCry14Ab</i>	<i>B. thuringiensis</i>	Biolistic method	Enhance resistance to soybean cyst nematodes	Jack	Kahn et al., 2021
	<i>GmSYP31A</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance resistance to soybean cyst nematodes	Williams82	Wang et al., 2021b
	<i>Hg-rps23</i> , <i>Hg-snb1</i> , and <i>Hg-cpn1</i>	<i>H. glycines</i>	<i>A. tumefaciens</i>	Enhance resistance to soybean cyst nematodes	Williams82	Zhang et al., 2022c
	<i>BtCry8</i> -like gene	<i>B. thuringiensis</i>	<i>A. tumefaciens</i>	Result in resistance to <i>Holotrichia parallela</i>	Jinong28	Qin et al., 2019
	<i>BtCry1la5</i>	<i>B. thuringiensis</i>	<i>A. tumefaciens</i>	Result in resistance to <i>Spodoptera littoralis</i>	Giza21 and Giza111	Moghaieb et al., 2019
Virus resistance	The coat protein gene of MYMIV	<i>Mung bean yellow mosaic India virus</i> (MYMIV)	<i>A. tumefaciens</i>	Result in resistance to yellow mosaic viruses	JS335	Kumari et al., 2018
Fungal disease resistance	SMV P3 cistron fragment (2,529–2,830 nt)	SMV SC3	<i>A. tumefaciens</i>	Enhance resistance to multiple Potyvirus strains and isolates	Shennong9 and Williams82	Yang X. et al., 2018
	<i>Gmelf4E</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in resistance to multiple potyvirids	Tianlong1	Gao et al., 2020
	The <i>AC2</i> gene	MYMIV	<i>A. tumefaciens</i>	Enhance MYMIV resistance	JS335	Ramesh et al., 2019
	<i>GmKR3</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in resistance to multiple viruses	Jack	Xun et al., 2019
	The protein kinase PBS1	TuMV	<i>A. tumefaciens</i>	Enhance potyvirus resistance	Williams82	Pottinger et al., 2020
	<i>GmVma12</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance SMV resistance	Tianlong1	Luan et al., 2020
	<i>GmST1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance resistance to soybean mosaic virus strains G2 and G3	Dongnong93–046	Zhao et al., 2021
	<i>GmNF-YC4-2</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in broad disease resistance for bacterial, viral, and fungal infections	Williams82	O'Conner et al., 2021
	<i>hrpZm</i>	<i>P. syringae</i>	<i>A. tumefaciens</i>	Enhance tolerance to Phytophthora root and stem rot caused by <i>P. sojae</i>	Williams82 and Shennong9	Du et al., 2018
	<i>AtPSS1</i>	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Result in resistance to <i>F. virguliforme</i>	Williams82	Wang et al., 2018a
	<i>GmCHI1A</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Result in resistance to <i>P. sojae</i>	Williams82 (carrying Rps 1k)	Zhou et al., 2018
	<i>GmPI4L</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in resistance to <i>P. sojae</i>	Dongnong50	Chen et al., 2019c
	<i>Hrf2</i>	<i>X. oryzaepv. oryzicola</i>	<i>A. tumefaciens</i>	Result in resistance to <i>P. sojae</i>	Shennong9	Niu et al., 2019
	<i>GmSnRK1.1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in resistance to <i>P. sojae</i>	Suinong10	Wang et al., 2019a
	<i>GmC4H1</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Result in resistance to <i>P. sojae</i>	Conrad	Yan et al., 2019
	<i>TaOXO</i>	<i>T. aestivum</i>	<i>A. tumefaciens</i>	Enhance resistance to sclerotinia stem rot	Williams82	Yang X. et al., 2019
	<i>GmMYB29A2</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Result in resistance to <i>P. sojae</i>	Harosoy, H63, Williams, and W82	Jahan et al., 2020
	<i>NmDef02</i>	<i>N. megalosiphon</i>	Biolistic method	Enhance resistance to soybean rust and anthracnose	DT-84	Soto et al., 2020
	<i>CmCH1</i>	<i>C. minitans</i>	<i>A. tumefaciens</i>	Enhanced resistance to <i>Sclerotinia sclerotiorum</i>	Williams82	Yang et al., 2020c
	<i>GmDR1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Result in broad spectrum immunity against fungal disease	Williams82	Ngaki et al., 2021
	<i>AtFOLT1</i>	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Enhance resistance to broad-spectrum disease	Williams82	Kambakam et al., 2021
	<i>NLR</i> gene	<i>G. max</i>	<i>O. haywardense</i>	Result in broad-spectrum resistance to <i>P. sojae</i>	93Y21	Wang et al., 2021c
	<i>GmTNL16</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance resistance to <i>P. sojae</i>	Williams	Li et al., 2022
	<i>GmNAC1</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance resistance to <i>P. sojae</i>	Tianlong1 and Suinong10	Yu et al., 2022

(Continued)

TABLE 2 | (Continued)

Target traits	Transgene	Source of gene	Delivery method	Effect on trait or function	Genotype	References
Herbicide tolerance	G10-EPSPS	<i>D. radiodurans</i>	<i>A. tumefaciens</i>	Result in glyphosate tolerance	Zhongdou32	Xiao et al., 2019
	G2-EPSPS and G10-EPSPS	<i>P. fluorescens</i>	<i>A. tumefaciens</i>	Result in glyphosate tolerance	Jack	Guo et al., 2020a
	Cytochrome P450 geneP450-N-Z1	<i>C. dactylon</i>	<i>A. tumefaciens</i>	Result in multiple herbicides tolerance	Tianlong1	Zheng et al., 2022
Drought tolerance	PgTIP1	<i>P. ginseng</i>	<i>A. tumefaciens</i>	Enhance both salt and drought tolerance	Hybrid strain 4076	An et al., 2018
	GmPIP2;9	<i>G. max</i>	<i>A. tumefaciens</i>	Increase drought tolerance	Williams82	Lu et al., 2018
	AtABF3	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	Kwangankong	Kim et al., 2018
	GmWRKY12	<i>G. max</i>	<i>A. rhizogenes</i>	Increase drought and salt tolerance	Williams82	Shi et al., 2018
	GmBIN2	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance tolerance to salt and drought	Dongnong50	Wang et al., 2018b
	GmBiP	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	Conquista	Coutinho et al., 2019
	AtYUCCA6	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	Kwangankong	Park et al., 2019
	GmWRKY54	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance drought tolerance	Williams82	Wei W. et al., 2019
	FvC5SD	<i>F. velutipes</i>	<i>A. tumefaciens</i>	Enhance drought stress tolerance	Shennong9	Zhang et al., 2019b
	GmNFYA5	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance drought tolerance	Williams82	Ma et al., 2020
	AtNCED3	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	BRS184	Molinari et al., 2020
	GmDREB2	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	DT84	Pham et al., 2020
	AtΔKinase	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Increase tolerance to water deficit stress	Williams82	Shanmugam et al., 2020
	HaHB4	<i>H. annuus</i>	<i>A. tumefaciens</i>	Enhance drought tolerance with yield reduced	Williams82	Ribichich et al., 2020
	GmNAC8	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	Tianlong1	Yang et al., 2020a
	GmbZIP2	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance tolerance to salt, drought, or cold condition	Williams82	Yang et al., 2020d
	GmbZIP15	<i>G. max</i>	<i>A. tumefaciens</i>	Decrease tolerance to drought and salt tolerance	C03-3	Zhang M. et al., 2020
	Gmgma-miR398c	<i>G. max</i>	<i>A. rhizogenes</i>	Negatively regulate drought tolerance	Williams82	Zhou et al., 2020
	GmNTF2B-1	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance drought tolerance	Williams82	Chen et al., 2021a
	GmMYB14	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance high-density yield and drought tolerance	Tianlong1	Chen et al., 2021b
	GmTGA15	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance drought tolerance	Williams82	Chen et al., 2021c
	GmPI-PLC7	<i>G. max</i>	<i>A. rhizogenes</i>	Increase drought and salt tolerance	Williams82	Chen et al., 2021d
	GmCIPK2	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	Williams82	Xu et al., 2021
	GsPOD40	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	PI342618B/DTP and Tianlong1	Aleem et al., 2022
	GmDREB1	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	P3	Chen et al., 2022
	sHSP26	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	Jinong18	Liu et al., 2022b
	GmDREB2	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought tolerance	BRS283	Marinho et al., 2022
	GmEF8	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance drought and heat tolerance	Williams82	Zhang et al., 2022a

(Continued)

TABLE 2 | (Continued)

Target traits	Transgene	Source of gene	Delivery method	Effect on trait or function	Genotype	References
Salt and other stress tolerance	<i>PgTIP1</i>	<i>P. ginseng</i>	<i>A. tumefaciens</i>	Enhance both salt and drought tolerance	Hybrid strain 4076	An et al., 2018
	<i>ZmGB1</i>	<i>Z. mays</i>	<i>A. tumefaciens</i>	Enhance tolerance to abiotic stress	\	Castiglioni et al., 2018
	<i>GmWRKY12</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase drought and salt tolerance	Williams82	Shi et al., 2018
	<i>AtXTH31</i>	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Enhance tolerance to flooding stress	Maverick	Song et al., 2018
	<i>GmBIN2</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance tolerance to salt and drought	Dongnong50	Wang et al., 2018b
	<i>MsWRKY11</i>	<i>M. sativa</i> (alfalfa)	<i>A. tumefaciens</i>	Enhance salt tolerance	Dongnong50	Wang et al., 2018c
	<i>GmHsp90A2</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase tolerance to heat stress	Qihuang22	Huang et al., 2019
	<i>AtAVP1</i> and <i>AtNHX1</i>	<i>A. thaliana</i>	<i>A. tumefaciens</i>	Increase salt tolerance	DT26	Nguyen N. T. et al., 2019
	<i>GmDREB-6</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance salt tolerance	DT84	Nguyen Q. H. et al., 2019
	<i>GsCLC-c2</i>	<i>G. soja</i>	<i>A. tumefaciens</i>	Enhance salt tolerance	N23674	Wei P. et al., 2019
	<i>GmERF135</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance salt tolerance	Tiefeng8	Zhao et al., 2019a
	<i>GmCDF1</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Negatively regulate salt tolerance	Kefeng1 and Nannong1138-2	Zhang et al., 2019c
	<i>GmSAP16</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance drought and salt tolerance	Williams82	Zhang et al., 2019d
	<i>J</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase salt tolerance	Huaxia3	Cheng et al., 2020
	<i>GsSnRK1</i>	<i>G. soja</i>	<i>A. tumefaciens</i>	Increase salt and alkaline stresses tolerance	Dongnong50	Feng et al., 2020
	<i>GmMYB68</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Increase salt and alkaline stresses tolerance	Williams82	He et al., 2020
	<i>Gs5PTase8</i>	<i>G. soja</i>	<i>A. rhizogenes</i>	Enhance salt tolerance	Mengjin1 and Union	Jia et al., 2020
	<i>GsAAE3</i>	<i>G. soja</i>	<i>A. rhizogenes</i>	Increase tolerance to Cd and Al stresses	BW69	Xian et al., 2020
	<i>GmbZIP2</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Enhance tolerance to salt, drought, or cold condition	Williams82	Yang et al., 2020d
	<i>GsJAZ2</i>	<i>G. soja</i> (G07256)	Biolistic method	Enhance tolerance to alkaline stress	HF55	Zhao et al., 2020
	<i>GmbZIP15</i>	<i>G. max</i>	<i>A. tumefaciens</i>	Decrease tolerance to drought and salt tolerance	C03-3	Zhang M. et al., 2020
	<i>GmPI-PLC7</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Increase drought and salt tolerance	Williams82	Chen et al., 2021d
	<i>AgGlpF</i>	<i>A. glaucus</i>	<i>A. tumefaciens</i>	Enhance salt tolerance	Williams82	Li et al., 2021a
	<i>GmNAC06</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance salt tolerance	Williams82	Li et al., 2021b
	<i>GsCLC-c2</i>	<i>G. soja</i>	<i>A. tumefaciens</i>	Enhance Cl ⁻ /salt tolerance	BB52	Liu et al., 2021
	<i>GsBET11a</i>	<i>G. soja</i>	<i>A. tumefaciens</i>	Enhance salt tolerance	G07256 and Dongnong50	Sun X. et al., 2021
	<i>GmNHX5</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance salt tolerance	Jidou-7	Sun T. et al., 2021
	<i>GmAKT1</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance salt tolerance	Dongnong50	Wang et al., 2021e
	<i>GmbHLH3</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance Cl ⁻ /salt tolerance	N23674	Liu et al., 2022c
	<i>GmEF8</i>	<i>G. max</i>	<i>A. rhizogenes</i>	Enhance drought and heat tolerance	Williams82	Zhang et al., 2022a

\ means not available.

(Pareddy et al., 2020) and 7 out of 8 genotypes (Zhao et al., 2019b). One important progress in these reports is that over 5% of TFs were obtained in nearly half of these genotypes. The second factor that affects genotype flexibility is the regeneration ability of donor genotypes, which restricts TFs for both *Agrobacterium*-mediated and biolistic transformations. Increasing the amenability of many soybean genotypes to regenerate may be conducted by either adding some specific chemicals in the culture medium described above or using plant regeneration factors or regeneration booster genes. Significant progress has been made to improve transformations from various tissue types using plant regeneration factors such as maize (*Zea mays*) morphogenic genes, *Baby boom* (BBM) and *Wuschel2* (WUS2) genes in maize plant (Lowe et al., 2016), and plant growth regulators such as *GROWTH-REGULATING FACTORS* (GRF) genes used in monocot and dicot species including soybean (Gordon-Kamm et al., 2019; Debernardi et al., 2020; Kong et al., 2020). Use of these genes significantly increased transformation frequency and reduced genotype obstacle for transformation, providing a good solution for genotype-inflexibility bottleneck in transformation of crops including soybean. For example, introducing *AtGRF5* and *GRF5* orthologs into soybean cells could improve regeneration and, hence, increase transformation TFs significantly (Kong et al., 2020). GRFs can also enhance shoot organogenesis and callus regeneration, which has been confirmed in dicots including sugar beet, canola, and sunflower. Meanwhile, somatic embryogenesis can be promoted using some genes introduced into explants in soybean, such as soybean orthologs of the *Arabidopsis* (*A. thaliana*) MADS box genes *AGAMOUS-Like15* (*GmAGL15*) and *GmAGL18*, which can also expand soybean genotypes suitable for transformation, especially for biolistic transformation (Zheng and Perry, 2014). Transformation bypass tissue culture such as *in planta* transformation is an alternative way to overcome genotype inflexibility in soybean (Liu et al., 2009; Mangena, 2019). Nanotechnology-based transformation can also be employed to overcome host range limitation including genotype inflexibility, and can simplify delivery way using pollen channel, and highly increase efficiency (Wang and Zhao, 2019). By integration of multiple-omics technologies, genes related to transformation efficiency should be discovered for increasing transformation efficiency. Use of the novel bacterium *O. haywardense* H1 may also increase the genotype scope for transformation, since it was claimed to be less genotype sensitive when it was used for soybean transformation (Cho et al., 2022).

Low Transformation Frequency

The average TF for varieties (genotypes) reported is lower than 5%, although improvements have been made by modifying the main factors described above (Table 1). Since the biolistic method tends to use an embryogenic callus as explant because of less chimerism compare to an embryo axis, TFs for biolistic transformation are dependent on the success of embryogenic callus induction for a specific genotype. Therefore, the main focus to improve TFs is to select genotypes that are amenable to embryogenic callus induction, or to stimulate a genotype to produce an embryogenic callus. As described above, the

regeneration booster provides a new way to induce an embryogenic callus without genotype limitation, which has been confirmed in monocot plants (Lowe et al., 2018; Gordon-Kamm et al., 2019; Debernardi et al., 2020). Enhancement of TFs for soybean *Agrobacterium*-mediated transformation is mainly achieved by improving regeneration rates of explants and increasing the susceptibility of explants to *Agrobacterium*. Half-seed explants have been the major choice, because these explants could provide more nutrition and less damage than cotyledonary nodes (Table 1). Continuously modifying MS-based culture medium composition (Murashige and Skoog, 1962), especially by addition of chemicals discovered through the study of omics, has played a big role in TF improvement, and has been summarized in the section above. Combinations of many factors have promoted the TFs of soybean transformation (Table 1). More efforts should be made to increase the average TFs close to that of other major crops. Again, the morphogenic genes including GRFs described above may play an important role in enhancing soybean transformation frequency.

Chimerism in T₀ Transgenic Plant

Chimerism in legume transformation is fairly common, which causes non-transmission of transgenes to subsequent generations either completely or at a lower ratio expected by Mendelian genetics. Therefore, minimizing chimerism in transgenic plants is required to obtain transmission of transgenes to the T₁ generation. In soybean, *Agrobacterium*-mediated transformation of cotyledonary nodes by organogenesis has been extensively conducted for transgenic production in research and commercial product development (Barwale et al., 1986; Homrich et al., 2012; Yamada et al., 2012; Lee et al., 2013; Mariashibu et al., 2013; Mangena et al., 2017). Plant regeneration by organogenesis with an explant containing an embryo axis may be the main cause, since shoots regenerated from soybean shoot tips were derived from 3 superimposed cellular layers (L1, L2, and L3) in the original apical meristem (Christou, 1990; Christou and McCabe, 1992). Transformed cells existed primarily in the L1 and L2 layers but not in the L3 layer of the apical meristems of regenerated shoots, indicating possible escape in the regenerated shoots during transformation, and this chimerism has been confirmed (McCabe et al., 1988; Sato et al., 1993). Currently, the chimerism in transgenic soybean is still a major concern in the research community, and inheritance study has been always an important part in transformation protocol development (Pareddy et al., 2020). Improvement for reducing escapes or chimeric rate has been made when strict select stringency was used, especially some new selectable markers/reagents such as AHAS/imazapyr (Aragão et al., 2000; Rech et al., 2008), EPSPS/glyphosate (Martinell et al., 1999; Guo et al., 2015, 2020a; Soto et al., 2017), and AADA/spectinomycin (Martinell et al., 2002). Meanwhile, the modified protocols made use of specific explants, such as somatic embryogenic calli, to reduce the chance of infection with cells at the late development stage, and combined proper selection of chemical agents with high stringency to decrease escape rate dramatically, which led to more than 90% T₀ transgenic plants transmitting their transgenes into T₁ generation (Soto et al., 2017; Chennareddy et al., 2018; Guo et al., 2020a). Therefore,

TABLE 3 | List of soybean genes edited for functional genetics study and trait improvement using genome editing technology.

Trait	Gene/targeting location	GE platform	Delivery method	Edited events	Editing outcomes	References
Yield						
Plant architecture	<i>GmLHY1a</i> (Glyma.16G017400), <i>GmLHY1b</i> (Glyma.07G048500), <i>GmLHY2a</i> (Glyma.19G260900), and <i>GmLHY2b</i> (Glyma.03G261800)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Cheng et al., 2019
	<i>GmSPL9a</i> (Glyma.02G177500), <i>GmSPL9b</i> (Glyma.09G113800), <i>GmSPL9c</i> (Glyma.03G143100), and <i>GmSPL9d</i> (Glyma.19G146000)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Bao et al., 2019
	<i>GmAP1a</i> (Glyma.16G091300), <i>GmAP1b</i> (Glyma.08G269800), <i>GmAP1c</i> (Glyma.01G064200), and <i>GmAP1d</i> (Glyma.02G121600)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Chen et al., 2020d
Seed weight and organ size	<i>GmPPD1</i> (Glyma.10G244400) and <i>GmPPD2</i> (Glyma.20G150000)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Kanazashi et al., 2018
	<i>GmSWEET10a</i> (Glyma.15G049200) and <i>GmSWEET10b</i> (Glyma.08G183500)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Wang et al., 2020c
	<i>GmKIX8-1</i> (Glyma.17G112800)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Nguyen et al., 2021
Seed number	<i>GmJAG1</i> (Glyma.20G25000) and <i>GmJAG2</i> (Glyma.10G42020)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Cai et al., 2021
Photoperiod	<i>GmFT2a</i> (Glyma.16G26660)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Cai et al., 2018
	<i>GmE1</i> (Glyma.06G207800)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Han et al., 2019
	<i>GmFT2a</i> (Glyma.16G26660) and <i>GmFT5a</i> (Glyma.16G04830)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Cai et al., 2020b
	<i>GmFT2a</i> (Glyma.16G150700) and <i>GmFT4</i> (Glyma.08G363100)	BE base editor	<i>A. tumefaciens</i>	Whole plant	Base editing	Cai et al., 2020a
	<i>GmFT2b</i> (Glyma.16G26690)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Chen et al., 2020c
	<i>GmAP1a</i> (Glyma.16G091300), <i>GmAP1b</i> (Glyma.08G269800), <i>GmAP1c</i> (Glyma.01G064200), and <i>GmAP1d</i> (Glyma.02G121600)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Chen et al., 2020d
	<i>GmPRR37</i> (Glyma.12G073900)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Wang et al., 2020b
	<i>GmLUX1</i> (Glyma.12G060200) and <i>GmLUX2</i> (Glyma.11G136600)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Bu et al., 2021
	<i>GmLNK2a</i> (Glyma.04G141400), <i>GmLNK2b</i> (Glyma.11G154700), <i>GmLNK2c</i> (Glyma.13G199300), and <i>GmLNK2d</i> (Glyma.15G237600)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Li et al., 2021c
Nutrition and quality						
Storage protein	<i>Glyma.20G148400</i> , <i>Glyma.20G146200</i> , <i>Glyma.10G246300</i> , <i>Glyma.20G148200</i> , <i>Glyma.10G037100</i> , <i>Glyma.03G163500</i> , <i>Glyma.19G164900</i> , <i>Glyma.13G123500</i> , and <i>Glyma.19G164800</i>	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Li et al., 2019
Seed oil	<i>GmFAD2-1A</i> (Glyma.10G278000) and <i>GmFAD2-1B</i> (Glyma.20G111000)	TALENs	<i>A. rhizogenes</i> and <i>disarmed A. rhizogenes</i>	Hairy root and whole plant	Knockout (multiplex)	Haun et al., 2014
	<i>GmFAD2-2</i>	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	al Amin et al., 2019
	<i>GmFAD2-1A</i> (Glyma.10G278000)	ZFNs	Biolytic method	Whole plant	Knock in (NHEJ)	Bonawitz et al., 2019
	<i>GmFAD2-1A</i> (Glyma.10G278000) and <i>GmFAD2-1B</i> (Glyma.20G111000)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Do et al., 2019
	<i>GmGOLS1A</i> (Glyma.03G222000) and <i>GmGOLS1B</i> (Glyma.19G219100)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Le et al., 2020
	<i>GmFAD2-1A</i> (Glyma.10G278000) and <i>GmFAD2-2A</i> (Glyma.19G147300)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Wu et al., 2020
	<i>GmFATB1a</i> (Glyma.05G012300) and <i>GmFATB1b</i> (Glyma.17G012400)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Ma J. et al., 2021
	<i>Glyma.15G117700</i>	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Qu et al., 2021
	<i>GmLx1</i> (Glyma.13G347600), <i>GmLx2</i> (Glyma.13G347500), and <i>GmLx3</i> (Glyma.15G026300)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Wang et al., 2020a

(Continued)

TABLE 3 | (Continued)

Trait	Gene/targeting location	GE platform	Delivery method	Edited events	Editing outcomes	References
Disease resistance						
Cyst nematode resistance	<i>GmSyn12</i> (Glyma.12G194800), <i>GmSyn13</i> (Glyma.13G307600), <i>GmSyn16</i> (Glyma.16G154200), and <i>GmSyn02</i> (Glyma.02G072900)	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Dong et al., 2020
	<i>Rps1</i> families (Glyma.03G034400, Glyma.03G0034800, Glyma.03G039200, Glyma.03G039500, Glyma.03G037100, Glyma.03G037300, Glyma.03G037400, Glyma.03G037400, Glyma.03G037000, Glyma.03G034500, Glyma.03G039300, Glyma.03G045700, Glyma.03G043600, Glyma.03G045300, Glyma.03G043000, Glyma.03G043500, Glyma.03G044000, Glyma.03G043200, Glyma.03G045000, Glyma.03G046500, Glyma.03G047000, Glyma.03G043900) and <i>Rpp1L</i> families (Glyma.18G281700, Glyma.18G281600, Glyma.18G281500, and Glyma.18G280300)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Nagy et al., 2021
	<i>Rhg1</i> -locus (Glyma.18G02270), <i>DELLA18</i> (Glyma.18G040000), and <i>DELLA11</i> (Glyma.11G216500)	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Dong and Hudson, 2022
Insect resistance	<i>GmUGT</i> (Glyma.07G110300)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout	Zhang et al., 2022b
Abiotic stress tolerance						
Drought tolerance	<i>GmLHY1a</i> (Glyma.16G017400), <i>GmLHY1b</i> (Glyma.07G048500), <i>GmLHY2a</i> (Glyma.19G260900), and <i>GmLHY2b</i> (Glyma.03G261800)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Wang et al., 2021a
Salt tolerance	<i>GmNAC06</i> (Glyma.06G21020)	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout	Li et al., 2021b
Nitrogen fixation						
	<i>GmNSP1a</i> (Glyma.07G039400) and <i>GmNSP1b</i> (Glyma.16G008200)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	He et al., 2021
Herbicide resistance	<i>GmALS1</i> (Glyma.04G37270.1), <i>GmALS2</i> (Glyma.06G17790.1), <i>GmALS3</i> (Glyma.13G31470.1), and <i>GmALS4</i> (Glyma.15G07860.1)	CRISPR/Cas9	Biolistic method	Whole plant	Knockin (HDR)	Li et al., 2015
Root nodulation	<i>GmRIC1</i> (Glyma.13G292300), <i>GmRIC2</i> (Glyma.06G284100), <i>GmRDN1-1</i> (Glyma.02G279600), <i>GmRDN1-2</i> (Glyma.14G035100), and <i>GmRDN1-3</i> (Glyma.20G040500)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Bai et al., 2020
	<i>GmSPL9d</i> (Glyma.19G146000) and <i>GmmiR156</i>	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Yun et al., 2022
Allergy reduction	<i>Gly m Bd 28K</i> (Glyma.U020300) and <i>Gly m Bd 30K</i> (Glyma.08G116300)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Sugano et al., 2020
GE platform adoption in soybean						
	<i>GmDCL1a</i> (Glyma.03G42290), <i>GmDCL1b</i> (Glyma.19G45060), <i>GmDCL4a</i> (Glyma.17G11240), <i>GmDCL4b</i> (Glyma.13G22450), <i>GmRDR6a</i> (Glyma.04G07150), <i>GmRDR6b</i> (Glyma.06G07250), and <i>GmHEN1a</i> (Glyma.08G08650)	ZFNs	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Curtin et al., 2011
	<i>GmDCL4a</i> (Glyma.17G11240) and <i>GmDCL4b</i> (Glyma.13G22450)	ZFNs	<i>A. rhizogenes</i>	Hairy root	Knockout	Sander et al., 2011
	<i>Bar</i> transgene, <i>GmFEI1</i> (Glyma.01G35390), <i>GmFEI2</i> (Glyma.09G34940), and <i>GmSHR</i>	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Cai et al., 2015
	<i>GmGS</i> (Glyma.18G04660 and Glyma.18G041100) and <i>GmCHI20</i> (Glyma.20G38560 and Glyma.20G241500)	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout	Michno et al., 2015
	<i>GFP</i> transgene, <i>Glyma07g14530</i> , <i>01gDDM1</i> (Glyma.11G38150), <i>11gDDM1</i> (Glyma.11G07220), <i>Glyma04g36150</i> , <i>Glyma06g18790</i> , <i>miR1509</i> , and <i>miR1514</i>	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout	Jacobs et al., 2015

(Continued)

TABLE 3 | (Continued)

Trait	Gene/targeting location	GE platform	Delivery method	Edited events	Editing outcomes	References
Transgene-free edited events	<i>Glyma.06G14180</i> , <i>Glyma.08G02290</i> , and <i>Glyma.12G37050</i>	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Sun et al., 2015
	<i>GmPDS11</i> (<i>Glyma.11G253000</i>) and <i>GmPDS18</i> (<i>Glyma.18G003900</i>)	TALENs	<i>A. tumefaciens</i>	Whole plant	Knockout	Du et al., 2016
	<i>FAD2-1A</i> (<i>Glyma.10G42470</i>) and <i>FAD2-1B</i> (<i>Glyma.20G24530</i>)	CRISPR/AsCpf1 or LpCpf1	Protoplast transfection	Protoplast	Knockout (RNP)	Kim et al., 2017
	<i>GmIPK1</i> (<i>Glyma.14G072200</i>) and <i>GmIPK2</i> (<i>Glyma.12G240900</i>) (STU and TCTU system*)	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Carrijo et al., 2021
	<i>Glyma.15G249000</i> and <i>Glyma.13G259100</i>	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Luo et al., 2021
	<i>GmPDS11g</i> (<i>Glyma.11g253000</i>) and <i>GmPDS18g</i> (<i>Glyma.18g003900</i>)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Lu and Tian, 2022
	Target sites DD38 and DD51	CRISPR/Cas9	<i>O. haywardense</i> H1-8		Targeted insertion	Kumar et al., 2022
	<i>Gly m Bd 30K</i> (<i>Glyma.08G116300</i>)	CRISPR/Cas9	Biolistic method	Whole plant	Knockout	Adachi et al., 2021
	<i>GmAGO7a</i> (<i>Glyma.01G053100</i>) and <i>GmAGO7b</i> (<i>Glyma.02G111600</i>)	CRISPR/Cas9	<i>A. rhizogenes</i> and <i>A. tumefaciens</i>	Hairy root and whole plant	Knockout (multiplex)	Zheng et al., 2020
	Targeted deletions of DNA fragments	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (4.5 kb in <i>GmFT2a</i>)	Cai et al., 2018
Growth of soybean trichomes	<i>GmCPR5</i> (<i>Glyma.06G145800</i>)	CRISPR/Cas9	Biolistic method	Whole plant	Knockout	Campbell et al., 2019
Fertility	<i>GmMs1</i> (<i>Glyma.13G114200</i>)	CRISPR/Cas9	Biolistic method	Whole plant	Knockout	Nadeem et al., 2021
miRNA pathway and small RNA processing	<i>GmMs1</i> (<i>Glyma.13G114200</i>)	CRISPR/Cas9	Biolistic method	Whole plant	Knockout	Jiang et al., 2021
	<i>GmDCL1a</i> (<i>Glyma.03G42290</i>), <i>GmDCL1b</i> (<i>Glyma.19G45060</i>), <i>GmDCL4a</i> (<i>Glyma.17G11240</i>), <i>GmDCL4b</i> (<i>Glyma.13G22450</i>), <i>GmRDR6a</i> (<i>Glyma.04G07150</i>), <i>GmRDR6b</i> (<i>Glyma.06G07250</i>), <i>GmHEN1a</i> (<i>Glyma.08G08650</i>), and <i>GFP</i> transgene	ZFNs	<i>A. rhizogenes</i>	Hairy root	Knockout	Curtin et al., 2011
	<i>GmDRB2a</i> (<i>Glyma.12G075700</i>), <i>GmDRB2b</i> (<i>Glyma.11G145900</i>), <i>GmDCL3a</i> (<i>Glyma.04G057400</i>), <i>GmHEN1a</i> (<i>Glyma.08G081600</i>), and <i>GmHEN1b</i> (<i>Glyma.05G126600</i>)	CRISPR/Cas9	<i>A. rhizogenes</i>	Hairy root	Knockout (multiplex)	Curtin et al., 2018
	<i>GmDCL2a</i> (<i>Glyma.09G025400</i>), <i>GmDCL2b</i> (<i>Glyma.09G025300</i>), and <i>GmDCL3a</i> (<i>Glyma.04G057400</i>)	TALENs	Disarmed <i>A. rhizogenes</i>	Whole plant	Knockout (multiplex)	Curtin et al., 2018
	<i>GmSWEET15a</i> (<i>Glyma.05G126600</i>) and <i>GmSWEET15b</i> (<i>Glyma.05G1266000</i>)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Wang et al., 2019b
Sucrose export related embryo development						
Circadian rhythmicity	<i>GmLCLa1</i> (<i>Glyma.16G01980</i>), <i>GmLCLa2</i> (<i>Glyma.07G05410</i>), <i>GmLCLb1</i> (<i>Glyma.03G42260</i>), and <i>GmLCLb2</i> (<i>Glyma.19G45030</i>)	CRISPR/Cas9	<i>A. tumefaciens</i>	Whole plant	Knockout (multiplex)	Wang et al., 2020e
Soybean knockout library	70 sgRNAs to target 102 genes	CRISPR/Cas9	<i>A. tumefaciens</i> (pooled)	Whole plant	Knockout (multiplex)	Bai et al., 2020

*STU, single transcriptional unit; SpCas9 and sgRNA are driven by only one promoter; and the two-component transcriptional unit (TCTU) in the conventional system, and SpCas9 and sgRNA are under the control of different promoters.

transgenic soybean with chimeric issues may due to insufficient selection that existed in various protocols.

Development of Transformation Method for New Breeding Technology

Genome editing is the recent advancement in genome engineering, which has revolutionized crop research and plant breeding. GE, through site-specific nucleases (SSNs), can precisely make changes in targeted genome sequence sites by disruption including insertion and deletion, base changes, sequence replacement, and sequence insertion. SSNs include zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and CRISPR/CAS. GE is a fast-developing technology that will potentially play an important role in genomics study and will create opportunities for rapid development of elite cultivars with desired traits. The development of soybean GE has been reviewed in Xu et al. (2020). Recent GE applications in soybean for trait improvement have been summarized in **Table 3**. For example, function analysis of photo period-related genes such as *LHY* homologs, *J* and *E1*, and *tof 16* (*Time of Flowering 16*) using GE technology showed that more than 80% accessions in low latitude harbor the mutations of *tof16* and *J*, which suggests that loss of functions of *Tof16* and *J* is the major genetic basis of soybean adaptation into tropics. Therefore, maturity and yield traits can be quantitatively improved by modulating the genetic complexity of various alleles of *LHY* homologs, *J*, and *E1* (Bu et al., 2021; Dong et al., 2021). The findings uncover the adaptation trajectory of soybean from its temperate origin to the tropics. Knockout of *GmJAG1*, which controls the number of seeds per pod (NSPP), increases by over 8% the yield of a Chinese variety, Huachun 6 (Cai et al., 2021). *GmMs1* KO events in soybean were created, which showed male sterility phenotype (Jiang et al., 2021; Nadeem et al., 2021). SCN-resistant mechanisms such as t-SNAREs binding Rhg1 α -SNAP (Dong et al., 2020) and WI12_{Rhg1} interacting with DELLAs (Dong and Hudson, 2022) were found using GE as a tool. Targeted chromosome cleavage by CRISPR/Cas9 can conceivably induce rearrangements and, thus, emergence of new resistance gene paralogs. CRISPR/Cas9-mediated chromosome rearrangements in nucleotide-binding-site-leucine-rich-repeat (NBS-LRR) gene families of soybean produced a new disease-resistant gene (Nagy et al., 2021). Raffinose family oligosaccharides (RFOs) are major soluble carbohydrates in soybean seeds that cannot be digested by humans and other monogastric animals. Double mutation events, knockouts in two soybean galactinol synthase (GOLS) genes, *GmGOLS1A* and its homeolog *GmGOLS1B*, showed a reduction in the total RFO content of soybean seeds from 64.7 to 41.95 mg/g dry weight, a 35.2% decrease (Le et al., 2020). This product improved the soybean nutrition quality. Two transcription systems were also tested in soybean including the single transcriptional unit (STU), SpCas9 and sgRNA are driven by only one promoter, and in the conventional system, the two-component transcriptional unit (TCTU), SpCas9, is under the control of a pol II promoter, and sgRNAs are under the control of a pol III promoter. The results showed that

the STU is more efficient (Carrijo et al., 2021). Cpfl (Cas12a) systems have also been established in soybean for GE (Duan et al., 2021; Kim and Choi, 2021). Meanwhile, different GE systems for soybean have been established using specific editing reagent delivery methods developed for soybean transformation, which produce transgene-free GE events either with the biolistic method (Adachi et al., 2021) and selectable marker-free GT systems by *O. haywardense* H1-8-mediated delivery (Kumar et al., 2022), or by organ-specific editing using an egg cell-specific promoter (Zheng et al., 2020). All these GE studies on soybean demonstrate that the ability to conduct genome editing directly depends on plant transformation technologies, since recovery of stable events with the target gene edited is normally based on available transformation systems including editing reagent delivery and edited event regeneration. GE has the potential to avoid many regulatory issues regarding transgenics if specific editing reagents are used. Based on the CRISPR/CAS system, gRNA in the form of *in vitro*-synthesized RNA molecule, together with Cas9 as DNA construct, can be stably integrated into the host genome and constitutively expressed, which might lead to a transgenic event for a GE event. This issue can be resolved by introduction of editing tools without genomic integration or transient expression. Transgene-free or DNA-free edited events in many crops can now be obtained either by delivering the RNA form of sgRNA and Cas9 or Cas9 protein (RNP) using the biolistic method, or by protoplast transfection (Chen et al., 2019a; Xu et al., 2020; Gao, 2021; Kim and Choi, 2021). Transgene-free events can also be recovered with the *Agrobacterium*-mediated method without selection (Liang et al., 2017). However, genotype flexibility limitation is a major issue for soybean GE in the biolistic method, and low TF for some elite varieties is the main hurdle in the *Agrobacterium*-mediated method.

CONCLUSION

As summarized above, development of soybean transformation protocols, which pose genotype-flexibility and relatively high efficiency and can easily be adapted in any laboratory, is still a main task for researchers. Reducing biological restrictions such as genotype dependence or tissue-specific and method restrictions will eventually lead to transformation automation and versatile and high throughput, which will facilitate the application of next-generation breeding technologies such as genome editing for soybean improvement. These goals may be achieved with fast progress in fundamental research to unravel basic biological process and genetic background, especially when more regeneration regulators such as morphogenic genes are identified. Transgenic soybean in which various genes can be manipulated will accelerate the validation of gene function in the context of complex gene networks at different plant developmental stages, which will accelerate the understanding of the mechanism of soybean cell regeneration, and it is beneficial for us to modify transformation protocols. New technologies like nanoparticle delivery also bring us hope to break through these barriers as well as the transformation bypass method.

AUTHOR CONTRIBUTIONS

HX and YG collected the materials. HX drew the figures and tables. YR wrote the manuscript. LQ and YR designed the article structure. All authors contributed to the article and approved the submitted version.

REFERENCES

- Abbasi, Z., Hooshyar, S., and Bagherieh-Najjar, M. B. (2016). Improvement of callus production and shoot regeneration using various organs of soybean (*Glycine max* L. Merr) by response surface methodology. *Vitro Cell. Dev. Pl.* 52, 537–545. doi: 10.1007/s11627-016-9778-1
- Adachi, K., Hirose, A., Kanazashi, Y., Hibara, M., Hirata, T., Mikami, M., et al. (2021). Site-directed mutagenesis by biolistic transformation efficiently generates inheritable mutations in a targeted locus in soybean somatic embryos and transgene-free descendants in the T₁ generation. *Transgenic Res.* 30, 77–89. doi: 10.1007/s11248-020-00229-4
- al Amin, N., Ahmad, N., Wu, N., Pu, X., Ma, T., Du, Y., et al. (2019). CRISPR-Cas9 mediated targeted disruption of FAD2–2 microsomal omega-6 desaturase in soybean (*Glycine max* L.). *BMC Biotechnol.* 19:9. doi: 10.1186/s12896-019-0501-2
- Aleem, M., Riaz, A., Raza, Q., Aleem, M., Aslam, M., Kong, K., et al. (2022). Genome-wide characterization and functional analysis of class III peroxidase gene family in soybean reveal regulatory roles of GsPOD40 in drought tolerance. *Genomics* 114, 45–60. doi: 10.1016/j.ygeno.2021.11.016
- An, J., Cheng, C., Hu, Z., Chen, H., Cai, W., and Yu, B. (2018). The *Panax ginseng* *PgTIP1* gene confers enhanced salt and drought tolerance to transgenic soybean plants by maintaining homeostasis of water, salt ions and ROS. *Environ. Exp. Bot.* 155, 45–55. doi: 10.1016/j.envexpbot.2018.06.025
- Anguraj Vadivel, A. K., McDowell, T., Renaud, J. B., and Dhaubhadel, S. (2021). A combinatorial action of GmMYB176 and GmbZIP5 controls isoflavonoid biosynthesis in soybean (*Glycine max*). *Commun. Biol.* 4:356. doi: 10.1038/s42003-021-01889-6
- Aragão, F. J. L., Sarokin, L., Vianna, G. R., and Rech, E. L. (2000). Selection of transgenic meristematic cells utilizing a herbicidal molecule results in the recovery of fertile transgenic soybean [*Glycine max* (L.) Merrill] plants at a high frequency. *Theor. Appl. Genet.* 101, 1–6. doi: 10.1007/s001220051441
- Arun, M., Chinnathambi, A., Subramanyam, K., Karthik, S., Sivanandhan, G., Thebora, J., et al. (2016). Involvement of exogenous polyamines enhances regeneration and *Agrobacterium*-mediated genetic transformation in half-seeds of soybean. *3 Biotech* 6:148. doi: 10.1007/s13205-016-0448-0
- Arun, M., Subramanyam, K., Mariashibu, T. S., Thebora, J., Shivanandhan, G., Manickavasagam, M., et al. (2015). Application of sonication in combination with vacuum infiltration enhances the *Agrobacterium*-mediated genetic transformation in Indian soybean cultivars. *Appl. Biochem. Biotechnol.* 175, 2266–2287. doi: 10.1007/s12010-014-1360-x
- Arya, H., Singh, M. B., and Bhalla, P. L. (2021). Overexpression of PIF4 affects plant morphology and accelerates reproductive phase transitions in soybean. *Food Energy Secur.* 10:e291. doi: 10.1002/fes3.291
- Bahramnejad, B., Naji, M., Bose, R., and Jha, S. (2019). A critical review on use of *Agrobacterium rhizogenes* and their associated binary vectors for plant transformation. *Biotech. Adv.* 37:107405. doi: 10.1016/j.biotechadv.2019.06.004
- Bai, M., Yuan, J., Kuang, H., Gong, P., Li, S., Zhang, Z., et al. (2020). Generation of a multiplex mutagenesis population via pooled CRISPR-Cas9 in soya bean. *Plant Biotechnol. J.* 18, 721–731. doi: 10.1111/pbi.13239
- Bao, A., Chen, H., Chen, L., Chen, S., Hao, Q., Guo, W., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of *GmSPL9* genes alters plant architecture in soybean. *BMC Plant Biol.* 19:131. doi: 10.1186/s12870-019-1746-6
- Barampura, S., and Zhang, Z. J. (2011). “Recent advances in plant transformation,” in *Plant Chromosome Engineering. Methods in Molecular Biology (Methods and Protocols)*, ed. J. A. Birchler (Totowa, NJ: Humana Press), 1–35.
- Barwale, U. B., Kerns, H. R., and Widholm Jack, M. (1986). Plant regeneration from callus cultures of several soybean genotypes via embryogenesis and organogenesis. *Planta* 167, 473–481. doi: 10.1007/BF00391223

ACKNOWLEDGMENTS

This study was supported by the National Transgenic Major Program of China (2016ZX08004001) and the Agricultural Science and Technology Innovation Program (ASTIP) of Chinese Academy of Agricultural Sciences.

- Beyer, E. (1979). Effect of silver ion, carbon dioxide, and oxygen on ethylene action and metabolism. *Plant Physiol.* 63, 169–173. doi: 10.1104/pp.63.1.169
- Bonawitz, N. D., Ainley, W. M., Itaya, A., Chennareddy, S. R., Cicak, T., Effinger, K., et al. (2019). Zinc finger nuclease-mediated targeting of multiple transgenes to an endogenous soybean genomic locus via non-homologous end joining. *Plant Biotechnol. J.* 17, 750–761. doi: 10.1111/pbi.13012
- Broothaerts, W., Mitchell, H. J., Weir, B., Kaines, S., Smith, L. M., Yang, W., et al. (2005). Gene transfer to plants by diverse species of bacteria. *Nature* 433, 629–633. doi: 10.1038/nature03309
- Bu, T., Lu, S., Wang, K., Dong, L., Li, S., Xie, Q., et al. (2021). A critical role of the soybean evening complex in the control of photoperiod sensitivity and adaptation. *Proc. Natl. Acad. Sci. U. S. A.* 118:e2010241118. doi: 10.1073/pnas.2010241118
- Cai, Y., Chen, L., Liu, X., Guo, C., Sun, S., Wu, C., et al. (2018). CRISPR/Cas9-mediated targeted mutagenesis of *GmFT2a* delays flowering time in soya bean. *Plant Biotechnol. J.* 16, 176–185. doi: 10.1111/pbi.12758
- Cai, Y., Chen, L., Liu, X., Sun, S., Wu, C., Jiang, B., et al. (2015). CRISPR/Cas9-mediated genome editing in soybean hairy roots. *PLoS One* 10:e0136064. doi: 10.1371/journal.pone.0136064
- Cai, Y., Wang, L., Chen, L., Wu, T., Liu, L., Sun, S., et al. (2020b). Mutagenesis of *GmFT2a* and *GmFT5a* mediated by CRISPR/Cas9 contributes to expanding the regional adaptability of soybean. *Plant Biotechnol. J.* 18, 298–309. doi: 10.1111/pbi.13199
- Cai, Y., Chen, L., Zhang, Y., Yuan, S., Su, Q., Sun, S., et al. (2020a). Target base editing in soybean using a modified CRISPR/Cas9 system. *Plant Biotechnol. J.* 18, 1996–1998. doi: 10.1111/pbi.13386
- Cai, Z., Xian, P., Cheng, Y., Ma, Q., Lian, T., Nian, H., et al. (2021). CRISPR/Cas9-mediated gene editing of *GmJAGGED1* increased yield in the low latitude soybean variety Huachun 6. *Plant Biotechnol. J.* 19, 1898–1900. doi: 10.1111/pbi.13673
- Campbell, B. W., Hoyle, J. W., Bucciarelli, B., Stec, A. O., Samac, D. A., Parrott, W. A., et al. (2019). Functional analysis and development of a CRISPR/Cas9 allelic series for a CPR5 ortholog necessary for proper growth of soybean trichomes. *Sci. Rep.* 9:14757. doi: 10.1038/s41598-019-51240-7
- Cao, D., Li, Y., Liu, B., Kong, F., and Tran, L. P. (2018). Adaptive mechanisms of soybean grown on salt-affected soils. *Land Degrad. Dev.* 29, 1054–1064. doi: 10.1002/ldr.2754
- Carrijo, J., Illa-Berenguer, E., LaFayette, P., Torres, N., Aragao, F. J. L., Parrott, W., et al. (2021). Two efficient CRISPR/Cas9 systems for gene editing in soybean. *Transgenic Res.* 30, 239–249. doi: 10.1007/s11248-021-00246-x
- Carvalho, C. H. S., Bohorova, N., Bordallo, P. N., Abreu, L. L., Valicente, F. H., Bressan, W., et al. (1997). Type II callus production and plant regeneration in tropical maize genotypes. *Plant Cell Rep.* 17, 73–76. doi: 10.1007/s002990050355
- Castiglioni, P., Bell, E., Lund, A., Rosenberg, A. F., Galligan, M., Hinchey, B. S., et al. (2018). Identification of GB1, a gene whose constitutive overexpression increases glycinebetaine content in maize and soybean. *Plant Direct* 2:e00040. doi: 10.1002/pld3.40
- Chee, P. P., Fober, K. A., and Slightom, J. L. (1989). Transformation of soybean (*Glycine max*) by infecting germinating seeds with *Agrobacterium tumefaciens*. *Plant Physiol.* 91, 1212–1218. doi: 10.1104/pp.91.3.1212
- Chen, K., Liu, W., Li, X., and Li, H. (2020b). Overexpression of GmGASA32 promoted soybean height by interacting with GmCDC25. *Plant Signal. Behav.* 16:1855017. doi: 10.1080/15592324.2020.1855017
- Chen, B., Zhang, G., Li, P., Yang, J., Guo, L., Benning, C., et al. (2020a). Multiple GmWRI1s are redundantly involved in seed filling and nodulation by regulating plastidic glycolysis, lipid biosynthesis and hormone signalling in soybean (*Glycine max*). *Plant Biotechnol. J.* 18, 155–171. doi: 10.1111/pbi.13183

- Chen, L., Cai, Y., Qu, M., Wang, L., Sun, H., Jiang, B., et al. (2020c). Soybean adaption to high-latitude regions is associated with natural variations of GmFT2b, an ortholog of FLOWERING LOCUS T. *Plant Cell Environ.* 43, 934–944. doi: 10.1111/pce.13695
- Chen, L., Nan, H., Kong, L., Yue, L., Yang, H., Zhao, Q., et al. (2020d). Soybean AP1 homologs control flowering time and plant height. *J. Integr. Agr.* 62, 1868–1879. doi: 10.1111/jipb.12988
- Chen, L., Yang, H., Fang, Y., Guo, W., Chen, H., Zhang, X., et al. (2021b). Overexpression of GmMYB14 improves high-density yield and drought tolerance of soybean through regulating plant architecture mediated by the brassinosteroid pathway. *Plant Biotechnol. J.* 19, 702–716. doi: 10.1111/pbi.13496
- Chen, K., Su, C., Tang, W., Zhou, Y., Xu, Z., Chen, J., et al. (2021a). Nuclear transport factor GmNTF2B-1 enhances soybean drought tolerance by interacting with oxidoreductase GmOXR17 to reduce reactive oxygen species content. *Plant J.* 107, 740–759. doi: 10.1111/tpj.15319
- Chen, Z., Fang, X., Yuan, X., Zhang, Y., Li, H., Zhou, Y., et al. (2021c). Overexpression of transcription factor GmTGA15 enhances drought tolerance in transgenic soybean hairy roots and *Arabidopsis* plants. *Agronomy* 11, 170. doi: 10.3390/agronomy11010170
- Chen, Z. F., Ru, J. N., Sun, G. Z., Du, Y., Chen, J., Zhou, Y. B., et al. (2021d). Genomic-wide analysis of the PLC family and detection of GmPI-PLC7 responses to drought and salt stresses in soybean. *Front. Plant Sci.* 12:631470. doi: 10.3389/fpls.2021.631470
- Chen, K., Tang, W., Zhou, Y., Chen, J., Xu, Z., Ma, R., et al. (2022). AP2/ERF transcription factor GmDREB1 confers drought tolerance in transgenic soybean by interacting with GmERFs. *Plant Physiol. Biochem.* 170, 287–295. doi: 10.1016/j.plaphy.2021.12.014
- Chen, K., Wang, Y., Zhang, R., Zhang, H., and Gao, C. (2019a). CRISPR/Cas genome editing and precision plant breeding in agriculture. *Ann. Rev. Plant Biol.* 70, 667–697. doi: 10.1146/annurev-arplant-050718-100049
- Chen, L., Qin, L., Zhou, L., Li, X., Chen, Z., Sun, L., et al. (2019b). A nodule-localized phosphate transporter GmPT7 plays an important role in enhancing symbiotic N₂ fixation and yield in soybean. *New Phytol.* 221, 2013–2025. doi: 10.1111/nph.15541
- Chen, X., Fang, X., Zhang, Y., Wang, X., Zhang, C., Yan, X., et al. (2019c). Overexpression of a soybean 4-coumaric acid: coenzyme A ligase (GmPI4L) enhances resistance to *Phytophthora sojae* in soybean. *Front. Plant Sci.* 46:304–313. doi: 10.1071/FP18111
- Chen, L., Cai, Y., Liu, X., Yao, W., Guo, C., Sun, S., et al. (2018b). Improvement of soybean *Agrobacterium*-mediated transformation efficiency by adding glutamine and asparagine into the culture media. *Int. J. Mol. Sci.* 19:3039. doi: 10.3390/ijms19103039
- Chen, L., Cai, Y., Liu, X., Guo, C., Sun, S., Wu, C., et al. (2018a). Soybean hairy roots produced in vitro by *Agrobacterium rhizogenes*-mediated transformation. *Crop J.* 6, 162–171. doi: 10.1016/j.cj.2017.08.006
- Chen, L., Chunyu, Z., Mingxia, F., Wenjuan, M., Meiming, C., Fengchun, C., et al. (2018c). GmIDL2a and GmIDL4a, encoding the inflorescence deficient in abscission-like protein, are involved in soybean cell wall degradation during lateral root emergence. *Int. J. Mol. Sci.* 19:2262.
- Cheng, M., Fry, J. E., Pang, S., Zhou, H., Hironaka, C. M., Duncan, D. R., et al. (1997). Genetic transformation of wheat mediated by *Agrobacterium tumefaciens*. *Plant Physiol.* 115, 971–980. doi: 10.1104/pp.115.3.971
- Cheng, Q., Dong, L., Su, T., Li, T., Gan, Z., Nan, H., et al. (2019). CRISPR/Cas9-mediated targeted mutagenesis of *GmLHY* genes alters plant height and internode length in soybean. *BMC Plant Biol.* 19:562. doi: 10.1186/s12870-019-2145-8
- Cheng, Q., Gan, Z., Wang, Y., Lu, S., Hou, Z., Li, H., et al. (2020). The soybean gene *J* contributes to salt stress tolerance by up-regulating salt-responsive genes. *Front. Plant Sci.* 11:272. doi: 10.3389/fpls.2020.00272
- Cheng, Y., Wang, X., Cao, L., Ji, J., Liu, T., and Duan, K. (2021). Highly efficient *Agrobacterium rhizogenes*-mediated hairy root transformation for gene functional and gene editing analysis in soybean. *Plant Methods* 17:73. doi: 10.1186/s13007-021-00778-7
- Chennareddy, S., Cicak, T., Mall, T., Effinger, K., Sardesai, N., Paredy, D., et al. (2018). Improved direct transformation via particle bombardment of split-immature embryo explants in soybean (*Glycine max*). *Plant Cell Tiss. Org.* 135, 23–35. doi: 10.1007/s11240-018-1440-7
- Cho, H. J., Farrand, S. K., Noel, G. R., and Widholm, J. M. (2000). High-efficiency induction of soybean hairy roots and propagation of the soybean cyst nematode. *Planta* 210, 195–204. doi: 10.1007/PL00008126
- Cho, H. J., Moy, Y., Rudnick, N. A., Klein, T. M., Yin, J., Bolar, J., et al. (2022). Development of an efficient marker-free soybean transformation method using the novel bacterium *Ochrobactrum haywardense* H1. *Plant Biotechnol. J.* 20, 977–990. doi: 10.1111/pbi.13777
- Cho, H. S., Lee, Y. J., Kim, H. J., Park, M.-Y., Yeom, W. W., Song, J. H., et al. (2021). Overexpression of *Arabidopsis thaliana* blue-light inhibitor of cryptochromes 1 gene alters plant architecture in soybean. *Plant Biotechnol. Rep.* 15, 459–469. doi: 10.1007/s11816-021-00693-2
- Chowrira, G. M., Akella, V., and Lurquin, P. F. (1995). Electroporation-mediated gene transfer into intact nodal meristems in planta. *Mol. Biotech.* 3, 17–23. doi: 10.1007/BF02821331
- Christey, M. C. (2001). Use of ri-mediated transformation for production of transgenic plants. *In Vitro Cell. Dev. Pl.* 37, 687–700. doi: 10.1007/s11627-001-0120-0
- Christou, P. (1990). Morphological description of transgenic soybean chimeras created by the delivery, integration and expression of foreign DNA using electric discharge particle acceleration. *Ann. Bot.* 66, 379–386. doi: 10.1093/oxfordjournals.aob.a088039
- Christou, P., McCabe, D., and Swain, W. (1988). Stable transformation of soybean callus by DNA-coated gold particles. *Plant Physiol.* 87, 671–674. doi: 10.1104/pp.87.3.671
- Christou, P., and McCabe, D. E. (1992). Prediction of germ-line transformation events in chimeric Ro transgenic soybean plantlets using tissue-specific expression patterns. *Plant J.* 2, 283–290. doi: 10.1111/j.1365-313X.1992.00283.x
- Clemente, T. E., LaVallee, B. J., Howe, A. R., Conner-Ward, D., Rozman, R. J., Hunter, P. E., et al. (2000). Progeny analysis of glyphosate selected transgenic soybeans derived from *Agrobacterium*-Mediated transformation. *Crop Sci.* 40, 797–803. doi: 10.2135/cropsci2000.403797x
- Collier, R., Fuchs, B., Walter, N., Kevin Lutke, W., and Taylor, C. G. (2005). Ex vitro composite plants: an inexpensive, rapid method for root biology. *Plant J.* 43, 449–457. doi: 10.1111/j.1365-313X.2005.02454.x
- Coutinho, F. S., Dos Santos, D. S., Lima, L. L., Vital, C. E., Santos, L. A., Pimenta, M. R., et al. (2019). Mechanism of the drought tolerance of a transgenic soybean overexpressing the molecular chaperone BiP. *Physiol. Mol. Biol. Plants* 25, 457–472. doi: 10.1007/s12298-019-00643-x
- Curtin, S. J., Xiong, Y., Michno, J. M., Campbell, B. W., Stec, A. O., Cermak, T., et al. (2018). CRISPR/Cas9 and TALENs generate heritable mutations for genes involved in small RNA processing of *Glycine max* and *Medicago truncatula*. *Plant Biotechnol. J.* 16, 1125–1137. doi: 10.1111/pbi.12857
- Curtin, S. J., Zhang, F., Sander, J. D., Haun, W. J., Starker, C., Baltes, N. J., et al. (2011). Targeted mutagenesis of duplicated genes in soybean with zinc-finger nucleases. *Plant Physiol.* 156, 466–473. doi: 10.1104/pp.111.172.981
- Dan, Y., Armstrong, C. L., Dong, J., Feng, X., Fry, J. E., Keithly, G. E., et al. (2009). Lipoic acid—an unique plant transformation enhancer. *In Vitro Cell. Dev. Biol. Plant* 45, 630–638. doi: 10.1007/s11627-009-9227-5
- Dan, Y., and Reichert, N. A. (1998). Organogenic regeneration of soybean from hypocotyl explants. *In Vitro Cell. Dev. Pl.* 34, 14–21. doi: 10.1007/BF02823117
- Dang, W., and Wei, Z. M. (2007). An optimized *Agrobacterium*-mediated transformation for soybean for expression of binary insect resistance genes. *Plant Sci.* 173, 381–389. doi: 10.1016/j.plantsci.2007.06.010
- Debernardi, J. M., Tricoli, D. M., Ercoli, M. F., Hayta, S., Ronald, P., Palatnik, J. F., et al. (2020). A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat. Biotech.* 38, 1274–1279. doi: 10.1038/s41587-020-0703-0
- Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol.* 16:225. doi: 10.1186/s12870-016-0906-1
- Dhir, S. K., Dhir, S., Savka, M. A., Belanger, F., Kriz, A. L., Farrand, S. K., et al. (1992). Regeneration of transgenic soybean (*Glycine max*) plants from electroporated protoplasts. *Plant Physiol.* 99, 81–88. doi: 10.1104/pp.99.1.81
- Do, P. T., Nguyen, C. X., Bui, H. T., Tran, L. T. N., Stacey, G., Gillman, J. D., et al. (2019). Demonstration of highly efficient dual gRNA CRISPR/Cas9 editing of the homeologous GmFAD2-1A and GmFAD2-1B genes to yield a high oleic,

- low linoleic and alpha-linolenic acid phenotype in soybean. *BMC Plant Biol.* 19:311. doi: 10.1186/s12870-019-1906-8
- Donaldson, P. A., and Simmonds, D. H. (2000). Susceptibility to *Agrobacterium tumefaciens* and cotyledonary node transformation in short-season soybean. *Plant Cell Rep.* 19, 478–484. doi: 10.1007/s002990050759
- Dong, J., and Hudson, M. E. (2022). *WI12^{Rhg1}* interacts with DELLAs and mediates soybean cyst nematode resistance through hormone pathways. *Plant Biotechnol. J.* 20, 283–296. doi: 10.1111/pbi.13709
- Dong, J., Zielinski, R. E., and Hudson, M. E. (2020). t-SNAREs bind the *Rhg1* α -SNAP and mediate soybean cyst nematode resistance. *Plant J.* 104, 318–331. doi: 10.1111/tj.14923
- Dong, L., Fang, C., Cheng, Q., Su, T., Kou, K., Kong, L., et al. (2021). Genetic basis and adaptation trajectory of soybean from its temperate origin to tropics. *Nat. Commun.* 12:5445. doi: 10.1038/s41467-021-25800-3
- Du, H., Zeng, X., Zhao, M., Cui, X., Wang, Q., Yang, H., et al. (2016). Efficient targeted mutagenesis in soybean by TALENs and CRISPR/Cas9. *J. Biotech.* 217, 90–97. doi: 10.1016/j.jbiotec.2015.11.005
- Du, Q., Yang, X., Zhang, J., Zhong, X., Kim, K. S., Yang, J., et al. (2018). Overexpression of the *Pseudomonas syringae* harpin-encoding gene *hrpZm* confers enhanced tolerance to *Phytophthora* root and stem rot in transgenic soybean. *Transgenic Res.* 27, 277–288. doi: 10.1007/s11248-018-0071-4
- Duan, K., Cheng, Y., Ji, J., Wang, C., Wei, Y., and Wang, Y. (2021). Large chromosomal segment deletions by CRISPR/LbCpf1-mediated multiplex gene editing in soybean. *J. Integr. Plant Biol.* 63, 1620–1631. doi: 10.1111/jipb.13158
- Dubey, V. K., Lee, U. G., Kwon, D. H., and Lee, S. H. (2017). Agroinfiltration-based expression of hairpin RNA in soybean plants for RNA interference against *Tetranychus urticae*. *Pestic. Biochem. Physiol.* 142, 53–58. doi: 10.1016/j.pestbp.2017.01.004
- Falco, S. C., Guida, T., Locke, M., Mauvais, J., Sanders, C., Ward, R. T., et al. (1995). Transgenic canola and soybean seeds with increased lysine. *Biotechnology* 13, 577–582. doi: 10.1038/nbt0695-577
- Fan, Y., Wang, X., Li, H., Liu, S., Jin, L., Lyu, Y., et al. (2020). Anthocyanin, a novel and user-friendly reporter for convenient, non-destructive, low cost, directly visual selection of transgenic hairy roots in the study of rhizobia-legume symbiosis. *Plant Methods* 16:94. doi: 10.1186/s13007-020-00638-w
- Faria, J. A. Q. A., Reis, P. A. B., Reis, M. T. B., Rosado, G. L., Pinheiro, G. L., Mendes, G. C., et al. (2011). The NAC domain-containing protein, GmNAC6, is a downstream component of the ER stress- and osmotic stress-induced NRP-mediated cell-death signaling pathway. *BMC Plant Biol.* 11:129. doi: 10.1186/1471-2229-11-129
- Feng, X., Feng, P., Yu, H., Yu, X., Sun, Q., Liu, S., et al. (2020). GsSnRK1 interplays with transcription factor GsERF7 from wild soybean to regulate soybean stress resistance. *Plant Cell Environ.* 43, 1192–1211. doi: 10.1111/pce.13726
- Fernandez, S., Michaux-Ferrière, N., and Coumans, M. F. (1999). The embryogenic response of immature embryo cultures of durum wheat (*Triticum durum* Desf.): histology and improvement by AgNO₃. *Plant Growth Regul.* 28, 147–155. doi: 10.1023/A:1006142504577
- Finer, J. J. (2016). Generation of transgenic soybean (*Glycine max*) via particle bombardment of embryogenic cultures. *Curr. Protoc. Plant Bio.* 1, 592–603. doi: 10.1002/cppb.20039
- Finer, J. J., and Larkin, K. M. (2008). “Genetic transformation of soybean using particle bombardment and SAAT approaches,” in *Handbook of New Technologies For Genetic Improvement of Legumes*, ed. K. PB (Boca Raton, USA: CRC Press/Taylor and Francis Group), 103–123.
- Finer, J. J., and McMullen, M. D. (1991). Transformation of soybean via particle bombardment of embryogenic suspension culture tissue. *In Vitro Cell. Dev. Pl.* 27, 175–182. doi: 10.2307/4292952
- Finer, J. J., and Nagasawa, A. (1988). Development of an embryogenic suspension culture of soybean (*Glycine max* Merrill.). *Plant Cell Tiss. Org.* 15, 125–136. doi: 10.1007/BF00035754
- Fu, M., Sun, J., Li, X., Guan, Y., and Xie, F. (2022). Asymmetric redundancy of soybean Nodule Inception (NIN) genes in root nodule symbiosis. *Plant Physiol.* 188, 477–489. doi: 10.1093/plphys/kiab473
- Gao, C. (2021). Genome engineering for crop improvement and future agriculture. *Cell* 184, 1621–1635. doi: 10.1016/j.cell.2021.01.005
- Gao, L., Luo, J., Ding, X., Wang, T., Hu, T., Song, P., et al. (2020). Soybean RNA interference lines silenced for eIF4E show broad potyvirus resistance. *Mol. Plant Pathol.* 21, 303–317. doi: 10.1111/mpp.12897
- Ghidelli, C., Mateos, M., Rojas-Argudo, C., and Pérez-Gago, M. B. (2014). Effect of antioxidants on enzymatic browning of eggplant extract and fresh-cut tissue. *J. Food Process. Pres.* 38, 1501–1510. doi: 10.1111/jfpp.12109
- Gordon-Kamm, W., Sardesai, N., Arling, M., Lowe, K., Hoerster, G., Betts, S., et al. (2019). Using morphogenic genes to improve recovery and regeneration of transgenic plants. *Plants* 8, 38. doi: 10.3390/plants8020038
- Grimsley, N., Hohn, B., Hohn, T., and Walden, R. (1986). “Agroinfection,” an alternative route for viral infection of plants by using the Ti plasmid. *Proc. Natl. Acad. Sci. U. S. A.* 83, 3282–3286. doi: 10.1073/pnas.83.10.3282
- Guo, B., Guo, Y., Hong, H., Jin, L., Zhang, L., Chang, R. Z., et al. (2015). Co-expression of G2-EPSPS and glyphosate acetyltransferase GAT genes conferring high tolerance to glyphosate in soybean. *Front. Plant Sci.* 6:847. doi: 10.3389/fpls.2015.00847
- Guo, B. F., Hong, H. L., Han, J. N., Zhang, L. J., and Qiu, L. J. (2020a). Development and identification of glyphosate-tolerant transgenic soybean via direct selection with glyphosate. *J. Integr. Agr.* 19, 1186–1196. doi: 10.1016/S2095-3119(19)62747-4
- Guo, C., Liu, X., Chen, L., Cai, Y., Yao, W., Yuan, S., et al. (2020b). Elevated methionine content in soybean seed by overexpressing maize β -zein protein. *Oil Crop Sci.* 5, 11–16. doi: 10.1016/j.ocsci.2020.03.004
- Guo, W., Chen, L., Chen, H., Yang, H., You, Q., Bao, A., et al. (2020c). Overexpression of GmWRI1b in soybean stably improves plant architecture and associated yield parameters, and increases total seed oil production under field conditions. *Plant Biotechnol. J.* 18, 1639–1641. doi: 10.1111/pbi.13324
- Hada, A., Gupta, A. K., Jeevaraj, T., Manickavasagam, M., Ganapathi, A., Jolly, M., et al. (2014). Developing rapid and reliable regeneration system using cotyledonary-node method in Indian soybean genotypes (*Glycine max* L.). *Int. J. Inno. Res. Sci.* 3, 12678–12686.
- Hada, A., Krishnan, V., Mohamed Jaabir, M. S., Kumari, A., Jolly, M., Praveen, S., et al. (2018). Improved *Agrobacterium tumefaciens*-mediated transformation of soybean [*Glycine max* (L.) Merr.] following optimization of culture conditions and mechanical techniques. *In Vitro Cell. Dev. Pl.* 54, 672–688. doi: 10.1007/s11627-018-9944-8
- Hadi, M. Z., McMullen, M. D., and Finer, J. J. (1996). Transformation of 12 different plasmids into soybean via particle bombardment. *Plant Cell Rep.* 15, 500–505. doi: 10.1007/BF00232982
- Han, J., Guo, B., Guo, Y., Zhang, B., Wang, X., and Qiu, L. J. (2019). Creation of early flowering germplasm of soybean by CRISPR/Cas9 technology. *Front. Plant Sci.* 10:1446. doi: 10.3389/fpls.2019.01446
- Han, X., Wang, D., and Song, G. Q. (2021). Expression of a maize SOC1 gene enhances soybean yield potential through modulating plant growth and flowering. *Sci. Rep.* 11:12758. doi: 10.1038/s41598-021-92215-x
- Hao, Y., Fan, X., Guo, H., Yao, Y., Ren, G., Lv, X., et al. (2020). Overexpression of the bioactive lunasin peptide in soybean and evaluation of its anti-inflammatory and anti-cancer activities invitro. *J. Biosci. Bioeng.* 129, 395–404. doi: 10.1016/j.jbiosc.2019.11.001
- Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. *Plant Biotechnol. J.* 12, 934–940. doi: 10.1111/pbi.12201
- Hayashi, M., Saeki, Y., Haga, M., Harada, K., Kouchi, H., and Umehara, Y. (2012). Rj (rj) genes involved in nitrogen-fixing root nodule formation in soybean. *Breeding Sci.* 61, 544–553. doi: 10.1270/jsbbs.61.544
- He, C., Gao, H., Wang, H., Guo, Y., He, M., Peng, Y., et al. (2021). GSK3-mediated stress signaling inhibits legume-rhizobium symbiosis by phosphorylating GmNSP1 in soybean. *Mol. Plant* 14, 488–502. doi: 10.1016/j.molp.2020.12.015
- He, Y., Dong, Y., Yang, X., Guo, D., Qian, X., Yan, F., et al. (2020). Functional activation of a novel R2R3-MYB protein gene, *GmMYB68*, confers salt-alkali resistance in soybean (*Glycine max* L.). *Genome* 63, 13–26. doi: 10.1139/gen-2018-0132
- Hernandez-Garcia, C. M., Martinelli, A. P., Bouchard, R. A., and Finer, J. J. (2009). A soybean (*Glycine max*) polyubiquitin promoter gives strong constitutive expression in transgenic soybean. *Plant Cell Rep.* 28, 837–849. doi: 10.1007/s00299-009-0681-7
- Hinchee, M. A. W., Connor-Ward, D. V., Newell, C. A., McDonnell, R. E., Sato, S. J., Gasser, C. S., et al. (1988). Production of transgenic soybean plants using *Agrobacterium*-mediated DNA transfer. *Biotechnology* 6, 915–922. doi: 10.1038/nbt0888-915

- Homrich, M. S., Wiebke-Strohm, B., Weber, R. L., and Bodanese-Zanettini, M. H. (2012). Soybean genetic transformation: a valuable tool for the functional study of genes and the production of agronomically improved plants. *Genet. Mol. Biol.* 35, 998–1010. doi: 10.1590/s1415-47572012000600015
- Hong, H., Zhang, H., Olhoft, P., Hill, S., Wiley, H., Toren, E., et al. (2007). Organogenic callus as the target for plant regeneration and transformation via *Agrobacterium* in soybean (*Glycine max* (L.) Merr.). *In Vitro Cell. Dev. Pl.* 43, 558–568. doi: 10.1007/s11627-007-9066-1
- Hu, C. Y., and Wang, L. (1999). In planta soybean transformation technologies developed in China: procedure, confirmation and field performance. *In Vitro Cell. Dev. Pl.* 35, 417–420. doi: 10.2307/4293277
- Huang, Y., Xuan, H., Yang, C., Guo, N., Wang, H., Zhao, J., et al. (2019). GmHsp90A2 is involved in soybean heat stress as a positive regulator. *Plant Sci.* 285, 26–33. doi: 10.1016/j.plantsci.2019.04.016
- Hwang, H. H., Yu, M., and Lai, E. M. (2017). *Agrobacterium*-mediated plant transformation: biology and applications. *Arabidopsis Book* 15:e0186. doi: 10.1199/tab.0186
- Islam, N., Islam, T., Hossain, M., Bhattacharjee, B., Hossain, M., and Islam, S. M. S. (2017). Embryogenic callus induction and efficient plant regeneration in three varieties of soybean (*Glycine max*). *Plant Tiss. Cul. Biotechnol.* 27, 41–50. doi: 10.3329/ptcb.v27i1.35011
- Ismagul, A., Yang, N., Maltseva, E., Iskakova, G., Mazonka, I., Skiba, Y., et al. (2018). A biolistic method for high-throughput production of transgenic wheat plants with single gene insertions. *BMC Plant Biol.* 18:135. doi: 10.1186/s12870-018-1326-1
- Itaya, A., Zheng, S., and Simmonds, D. (2018). Establishment of neomycin phosphotransferase II (nptII) selection for transformation of soybean somatic embryogenic cultures. *In Vitro Cell. Dev. Pl.* 54, 184–194. doi: 10.1007/s11627-017-9875-9
- Iwabuchi, K., Shimada, T. L., Yamada, T., and Hara-Nishimura, I. (2020). A space-saving visual screening method. *Glycine max* FAST, for generating transgenic soybean. *Plant Signal. Behav.* 15:1722911. doi: 10.1080/15592324.2020.1722911
- Jacobs, T. B., LaFayette, P. R., Schmitz, R. J., and Parrott, W. A. (2015). Targeted genome modifications in soybean with CRISPR/Cas9. *BMC Biotechnol.* 15:16. doi: 10.1186/s12896-015-0131-2
- Jahan, M. A., Harris, B., Lowery, M., Infante, A. M., Percifield, R. J., and Kovinich, N. (2020). Glyceollin transcription factor GmMYB29A2 regulates soybean resistance to *Phytophthora sojae*. *Plant Physiol.* 183, 530–546. doi: 10.1104/pp.19.01293
- Jia, Q., Sun, S., Kong, D., Song, J., Wu, L., Yan, Z., et al. (2020). Ectopic expression of *Gs5PTase8*, a soybean inositol polyphosphate 5-phosphatase, enhances salt tolerance in plants. *Int. J. Mol. Sci.* 21, 1023. doi: 10.3390/ijms21031023
- Jia, Y., Yao, X., Zhao, M., Zhao, Q., Du, Y., Yu, C., et al. (2015). Comparison of soybean transformation efficiency and plant factors affecting transformation during the *Agrobacterium* infection process. *Int. J. Mol. Sci.* 16, 18522–18543. doi: 10.3390/ijms160818522
- Jiang, B., Chen, L., Yang, C., Wu, T., Yuan, S., Wu, C., et al. (2021). The cloning and CRISPR/Cas9-mediated mutagenesis of a male sterility gene MS1 of soybean. *Plant Biotechnol. J.* 19, 1098–1100. doi: 10.1111/pbi.13601
- Jing, G., Tang, D., Yao, Y., Su, Y., Shen, Y., Bai, Y., et al. (2021). Seed specifically over-expressing *DGAT2A* enhances oil and linoleic acid contents in soybean seeds. *Biochem. Biophys. Res. Commun.* 568, 143–150. doi: 10.1016/j.bbrc.2021.06.087
- Jones, S. I., Hunt, M. R., and Vodkin, L. O. (2020). An embryo lethal transgenic line manifests global expression changes and elevated protein/oil ratios in heterozygous soybean plants. *PLoS One* 15:e0233721. doi: 10.1371/journal.pone.0233721
- Joyner, E. Y., Boykin, L. S., and Lodhi, M. A. (2010). Callus induction and organogenesis in soybean [*Glycine max* (L.) Merr.] cv. pyramid from mature cotyledons and embryos. *Open Plant Sci. J.* 4, 18–21. doi: 10.2174/1874294701004010018
- Kahn, T. W., Duck, N. B., McCarville, M. T., Schouten, L. C., Schweri, K., Zaitseva, J., et al. (2021). A *Bacillus thuringiensis* Cry protein controls soybean cyst nematode in transgenic soybean plants. *Nat. Commun.* 12:3380. doi: 10.1038/s41467-021-23743-3
- Kalbande, B. B., and Patil, A. S. (2016). Plant tissue culture independent *Agrobacterium tumefaciens* mediated In-planta transformation strategy for upland cotton (*Gossypium hirsutum*). *J. Genet. Eng. Biotechnol.* 14, 9–18. doi: 10.1016/j.jgeb.2016.05.003
- Kambakam, S., Ngaki, M. N., Sahu, B. B., Kandel, D. R., Singh, P., Sumit, R., et al. (2021). Arabidopsis non-host resistance *PSS30* gene enhances broad-spectrum disease resistance in the soybean cultivar Williams 82. *Plant J.* 107, 1432–1446. doi: 10.1111/tpj.15392
- Kanai, M., Yamada, T., Hayashi, M., Mano, S., and Nishimura, M. (2019). Soybean (*Glycine max* L.) triacylglycerol lipase GmSDP1 regulates the quality and quantity of seed oil. *Sci. Rep.* 9:8924. doi: 10.1038/s41598-019-45331-8
- Kanazashi, Y., Hirose, A., Takahashi, I., Mikami, M., Endo, M., Hirose, S., et al. (2018). Simultaneous site-directed mutagenesis of duplicated loci in soybean using a single guide RNA. *Plant Cell Rep.* 37, 553–563. doi: 10.1007/s00299-018-2251-3
- Karthik, S., Pavan, G., and Manickavasagam, M. (2020). Nitric oxide donor regulates *Agrobacterium*-mediated genetic transformation efficiency in soybean [*Glycine max* (L.) Merrill]. *Plant Cell Tiss. Org.* 141, 655–660. doi: 10.1007/s11240-020-01808-3
- Karthik, S., Tuteja, N., Ganapathi, A., and Manickavasagam, M. (2019). Pea p68, a DEAD-box helicase, enhances salt tolerance in marker-free transgenic plants of soybean [*Glycine max* (L.) Merrill]. *3 Biotech* 9:10. doi: 10.1007/s13205-018-1553-z
- Kereszt, A., Li, D., Indrasumunar, A., Nguyen, C. D., Nontachaiyapoom, S., Kinkema, M., et al. (2007). *Agrobacterium rhizogenes*-mediated transformation of soybean to study root biology. *Nat. Protoc.* 2, 948–952. doi: 10.1038/nprot.2007.141
- Kidokoro, S., Watanabe, K., Ohori, T., Moriwaki, T., Maruyama, K., Mizoi, J., et al. (2015). Soybean DREB1/CBF-type transcription factors function in heat and drought as well as cold stress-responsive gene expression. *Plant J.* 81, 505–518. doi: 10.1111/tpj.12746
- Kim, H., and Choi, J. (2021). A robust and practical CRISPR/crRNA screening system for soybean cultivar editing using LbCpf1 ribonucleoproteins. *Plant Cell Rep.* 40, 1059–1070. doi: 10.1007/s00299-020-02597-x
- Kim, H., Kim, S. T., Ryu, J., Kang, B. C., Kim, J. S., and Kim, S. G. (2017). CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.* 8:14406. doi: 10.1038/ncomms14406
- Kim, H. J., Cho, H. S., Pak, J. H., Kwon, T., Lee, J. H., Kim, D. H., et al. (2018). Confirmation of drought tolerance of ectopically expressed AtABF3 gene in soybean. *Mol. Cells* 41, 413–422. doi: 10.14348/molcells.2018.2254
- Kim, K. H., Lim, S., Kang, Y. J., Yoon, M. Y., Nam, M., Jun, T. H., et al. (2016). Optimization of a virus-induced gene silencing system with soybean yellow common mosaic virus for gene function studies in soybeans. *Plant Pathol. J.* 32, 112–122. doi: 10.5423/PPJ.OA.04.2015.0063
- Kim, W.-S., and Krishnan, H. B. (2019). Impact of co-expression of maize 11 and 18 kDa δ -zeins and 27 kDa γ -zein in transgenic soybeans on protein body structure and sulfur amino acid content. *Plant Sci.* 280, 340–347. doi: 10.1016/j.plantsci.2018.12.016
- Kim, W.-S., Sun-Hyung, J., Oehrle, N. W., Jez, J. M., and Krishnan, H. B. (2020). Overexpression of ATP sulfurylase improves the sulfur amino acid content, enhances the accumulation of Bowman-Birk protease inhibitor and suppresses the accumulation of the β -subunit of β -conglycinin in soybean seeds. *Sci. Rep.* 10:14989. doi: 10.1038/s41598-020-72134-z
- King, J. L., Finer, J. J., and McHale, L. K. (2015). Development and optimization of agroinfiltration for soybean. *Plant Cell Rep.* 34, 133–140. doi: 10.1007/s00299-014-1694-4
- Klein, T. M., Wolf, E. D., Wu, R., and Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* 327, 70–73. doi: 10.1038/327070a0
- Ko, T. S., Lee, S., Krasnyanski, S., and Korban, S. S. (2003). Two critical factors are required for efficient transformation of multiple soybean cultivars: *agrobacterium* strain and orientation of immature cotyledonary explant. *Theor. Appl. Genet.* 107, 439–447. doi: 10.1007/s00122-003-1264-6
- Kong, J., Martin-Ortigosa, S., Finer, J., Orchard, N., Gunadi, A., Batts, L. A., et al. (2020). Overexpression of the transcription factor GROWTH-REGULATING FACTOR5 improves transformation of dicot and monocot species. *Front. Plant Sci.* 11:572319. doi: 10.3389/fpls.2020.572319
- Kuma, K. M., Lopes-Caitar, V. S., Romero, C. C. T., Silva, S. M. H., Kuwahara, M. K., Carvalho, M. C. C. G., et al. (2015). A high efficient protocol for soybean root transformation by *Agrobacterium rhizogenes* and most stable reference genes for RT-qPCR analysis. *Plant Cell Rep.* 34, 1987–2000. doi: 10.1007/s00299-015-1845-2

- Kumar, A., Kumar, V., Krishnan, V., Hada, A., and Marathe, A. C. P. (2019). Seed targeted RNAi-mediated silencing of GmMIPS1 limits phytate accumulation and improves mineral bioavailability in soybean. *Sci. Rep.* 9:7744. doi: 10.1038/s41598-019-44255-7
- Kumar, S., Liu, Z.-B., Sanyour-Doyel, N., Lenderts, B., Worden, A., Anand, A., et al. (2022). Efficient gene targeting in soybean using *Ochrobactrum haywardense*-mediated delivery of a marker-free donor template. *Plant Physiol.* 00, 1–10. doi: 10.1093/plphys/kiac075
- Kumar, S. V., and Rajam, M. (2005). Polyamines enhance *Agrobacterium tumefaciens* vir gene induction and T-DNA transfer. *Plant Sci.* 168, 475–480. doi: 10.1016/j.plantsci.2004.09.018
- Kumari, A., Hada, A., Subramanyam, K., Thebora, J., Misra, S., Ganapathi, A., et al. (2018). RNAi-mediated resistance to yellow mosaic viruses in soybean targeting coat protein gene. *Acta Physiol. Plant* 40:32. doi: 10.1007/s11738-018-2608-9
- Kun, X., Zhang, X.-M., Fan, C.-M., Chen, F.-L., Zhu, J.-L., Zhang, S.-L., et al. (2017). A callus transformation system for gene functional studies in soybean. *J. Integr. Agr.* 16, 1913–1922. doi: 10.1016/S2095-3119(16)61621-0
- Le, H., Nguyen, N. H., Ta, D. T., Le, T. N. T., Bui, T. P., Le, N. T., et al. (2020). CRISPR/Cas9-mediated knockout of galactinol synthase-encoding genes reduces raffinose family oligosaccharide levels in soybean seeds. *Front. Plant Sci.* 11:612942. doi: 10.3389/fpls.2020.612942
- Lee, H., Park, S.-Y., and Zhang, Z. J. (2013). “An overview of genetic transformation of soybean,” in *A Comprehensive Survey of International Soybean Research - Genetics, Physiology, Agronomy and Nitrogen Relationships*, ed. J. E. Board (London, UK: IntechOpen).
- Li, C., Nguyen, V., Liu, J., Fu, W., Chen, C., Yu, K., et al. (2019). Mutagenesis of seed storage protein genes in soybean using CRISPR/Cas9. *BMC Res. Notes* 12:176. doi: 10.1186/s13104-019-4207-2
- Li, C., Zhang, H., Wang, X., and Liao, H. (2014). A comparison study of *Agrobacterium*-mediated transformation methods for root-specific promoter analysis in soybean. *Plant Cell Rep.* 33, 1921–1932. doi: 10.1007/s00299-014-1669-5
- Li, F., Ni, H., Yan, W., Xie, Y., Liu, X., Tan, X., et al. (2021a). Overexpression of an aquaporin protein from *Aspergillus glaucus* confers salt tolerance in transgenic soybean. *Transgenic Res.* 30, 727–737. doi: 10.1007/s11248-021-00280-9
- Li, M., Chen, R., Jiang, Q., Sun, X., Zhang, H., and Hu, Z. (2021b). GmNAC06, a NAC domain transcription factor enhances salt stress tolerance in soybean. *Plant Mol. Biol.* 105, 333–345. doi: 10.1007/s11103-020-01091-y
- Li, Z., Cheng, Q., Gan, Z., Hou, Z., Zhang, Y., Li, Y., et al. (2021c). Multiplex CRISPR/Cas9-mediated knockout of soybean LNK2 advances flowering time. *Crop J.* 9, 767–776. doi: 10.1016/j.cj.2020.09.005
- Li, J., Todd, T. C., Oakley, T. R., Lee, J., and Trick, H. N. (2010). Host-derived suppression of nematode reproductive and fitness genes decreases fecundity of *Heterodera glycines* Ichinohe. *Planta* 232, 775–785. doi: 10.1007/s00425-010-1209-7
- Li, L., Gao, W., Peng, Q., Zhou, B., Kong, Q., Ying, Y., et al. (2018). Two soybean bHLH factors regulate response to iron deficiency. *J. Integr. Plant Biol.* 60, 608–622. doi: 10.1111/jipb.12651
- Li, S., Cong, Y., Liu, Y., Wang, T., Shuai, Q., Chen, N., et al. (2017). Optimization of *Agrobacterium*-mediated transformation in soybean. *Front. Plant Sci.* 8:246. doi: 10.3389/fpls.2017.00246
- Li, Z., Liu, Z. B., Xing, A., Moon, B. P., Koellhoffer, J. P., Huang, L., et al. (2015). Cas9-guide RNA directed genome editing in soybean. *Plant Physiol.* 169, 960–970. doi: 10.1104/pp.15.00783
- Li, Z., Sushuang, D., Huidong, X., Xingxing, F., Ruidong, S., Jinming, Z., et al. (2022). A novel TIR-NBS-LRR gene regulates immune response to Phytophthora root rot in soybean. *Crop J.* doi: 10.1016/j.cj.2022.03.003
- Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., et al. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nat. Commun.* 8:14261. doi: 10.1038/ncomms14261
- Lin, W., Odell, J. T., and Schreiner, R. M. (1987). Soybean protoplast culture and direct gene uptake and expression by cultured soybean protoplasts. *Plant Physiol.* 84, 856–861. doi: 10.1104/pp.84.3.856
- Liu, C., Zhang, C., Fan, M., Ma, W., Chen, M., Cai, F., et al. (2018). *GmIDL2a* and *GmIDL4a*, encoding the inflorescence deficient in abscission-like protein, are involved in soybean cell wall degradation during lateral root emergence. *Int. J. Mol. Sci.* 19:2262. doi: 10.3390/ijms19082262
- Liu, W., Jiang, B., Ma, L., Zhang, S., Zhai, H., Xu, X., et al. (2018). Functional diversification of Flowering Locus T homologs in soybean: *GmFT1a* and *GmFT2a/5a* have opposite roles in controlling flowering and maturation. *New Phytol.* 217, 1335–1345. doi: 10.1111/nph.14884
- Liu, H. K., Yang, C., and Wei, Z. M. (2004). Efficient *Agrobacterium tumefaciens*-mediated transformation of soybeans using an embryonic tip regeneration system. *Planta* 219, 1042–1049. doi: 10.1007/s00425-004-1310-x
- Liu, J., Su, Q., An, L., and Yang, A. (2009). Transfer of a minimal linear marker-free and vector-free smGFP cassette into soybean via ovary-drip transformation. *Biotechnol. Lett.* 31, 295–303. doi: 10.1007/s10529-008-9851-x
- Liu, L. F., Le, G., Zhang, L. X., Cai, Y. P., Song, W. W., and Han, T. F. (2022a). Co-silencing E1 and its homologs in an extremely late-maturing soybean cultivar confers super-early maturity and adaptation to high-latitude short-season regions. *J. Integr. Agr.* 21, 326–335. doi: 10.1016/S2095-3119(20)63391-3
- Liu, S. Y., Liu, J. F., Zhang, Y. Z., Jiang, Y. S., Hu, S. W., Qu, J., et al. (2022b). Cloning of the soybean sHSP26 gene and analysis of its drought resistance. *Phyton* 97, 1465–1482. doi: 10.32604/phyton.2022.018836
- Liu, X., Pi, B., Du, Z., Yang, T., Gu, M., Sun, S., et al. (2022c). The transcription factor GmbHLH3 confers Cl⁻/salt tolerance to soybean by upregulating *GmCLC1* expression for maintenance of anion homeostasis. *Environ. Exp. Bot.* 194:104755. doi: 10.1016/j.envexpbot.2021.104755
- Liu, S., Wang, D., Mei, Y., Xia, T., Xu, W., Zhang, Y., et al. (2020). Overexpression of GmAAP6a enhances tolerance to low nitrogen and improves seed nitrogen status by optimizing amino acid partitioning in soybean. *Plant Biotechnol.* 18, 1749–1762. doi: 10.1111/pbi.13338
- Liu, S. J., Wei, Z. M., and Huang, J. Q. (2008). The effect of co-cultivation and selection parameters on *Agrobacterium*-mediated transformation of Chinese soybean varieties. *Plant Cell Rep.* 27, 489–498. doi: 10.1007/s00299-007-0475-8
- Liu, X., Liu, F., Zhang, L., Cheng, C., Wei, P., and Yu, B. (2021). *GsCLC-c2* from wild soybean confers chloride/salt tolerance to transgenic *Arabidopsis* and soybean composite plants by regulating anion homeostasis. *Physiol. Plantarum* 172, 1867–1879. doi: 10.1111/ppl.13396
- Lowe, K., La Rota, M., Hoerster, G., Hastings, C., Wang, N., Chamberlin, M., et al. (2018). Rapid genotype “independent” *Zea mays* L. (maize) transformation via direct somatic embryogenesis. *In Vitro Cell. Dev. Pl.* 54, 240–252. doi: 10.1007/s11627-018-9905-2
- Lowe, K., Wu, E., Wang, N., Hoerster, G., Hastings, C., Cho, M. J., et al. (2016). Morphogenic regulators *Baby boom* and *Wuschel* improve monocot transformation. *Plant Cell* 28, 1998–2015. doi: 10.1105/tpc.16.00124
- Lu, L., Dong, C., Liu, R., Zhou, B., Wang, C., and Shou, H. (2018). Roles of soybean plasma membrane intrinsic protein GmPIP2;9 in drought tolerance and seed development. *Front. Plant Sci.* 9:530. doi: 10.3389/fpls.2018.00530
- Lu, L., Wei, W., Li, Q.-T., Bian, X.-H., Lu, X., Hu, Y., et al. (2021). A transcriptional regulatory module controls lipid accumulation in soybean. *New Phytol.* 231, 661–678. doi: 10.1111/nph.17401
- Lu, Q. S. M., and Tian, L. (2022). An efficient and specific CRISPR-Cas9 genome editing system targeting soybean phytoene desaturase genes. *BMC Biotechnol.* 22:7. doi: 10.1186/s12896-022-00737-7
- Luan, H., Liao, W., Song, Y., Niu, H., and Zhi, H. (2020). Transgenic plant generated by RNAi-mediated knocking down of soybean Vma12 and soybean mosaic virus resistance evaluation. *AMB Express* 10, 1–10. doi: 10.1186/s13568-020-00997-6
- Luo, Y., Na, R., Nowak, J. S., Qiu, Y., Lu, Q. S., Yang, C., et al. (2021). Development of a Csy4-processed guide RNA delivery system with soybean-infecting virus ALSV for genome editing. *BMC Plant Biol.* 21:419. doi: 10.1186/s12870-021-03138-8
- Luth, D., Warnberg, K., and Wang, K. (2015). “Soybean [*Glycine max* (L.) Merr.]” in *Agrobacterium Protocols. Methods in Molecular Biology*, ed. K. Wang (New York, NY: Springer), 275–284.
- Ma, J., Sun, S., Whelan, J., and Shou, H. (2021). CRISPR/Cas9-mediated knockout of *GmFATB1* significantly reduced the amount of saturated fatty acids in soybean seeds. *Int. J. Mol. Sci.* 22:3877. doi: 10.3390/ijms22083877
- Ma, X., Xu, W., Liu, T., Chen, R., Zhu, H., Zhang, H., et al. (2021). Functional characterization of soybean (*Glycine max*) *DIRIGENT* genes reveals an important role of *GmDIR27* in the regulation of pod dehiscence. *Genomics* 113, 979–990. doi: 10.1016/j.ygeno.2020.10.033

- Ma, X. J., Yu, T. F., Li, X. H., Cao, X. Y., Ma, J., Chen, J., et al. (2020). Overexpression of GmNFYA5 confers drought tolerance to transgenic *Arabidopsis* and soybean plants. *BMC Plant Biol.* 20:123. doi: 10.1186/s12870-020-02337-z
- Mangena, P. (2019). A simplified in-planta genetic transformation in soybean. *Res. J. Biotech.* 14, 117–125.
- Mangena, P., Mokwala, P., and Nikolova, R. V. (2017). “Challenges of in vitro and in vivo *Agrobacterium*-mediated genetic transformation in soybean,” in *Soybean - The Basis of Yield, Biomass and Productivity*, ed. M. Kasai (London: IntechOpen), 75–94.
- Mariashibu, T. S., Subramanyam, K., Arun, M., Mayavan, S., Rajesh, M., Thebora, J., et al. (2013). Vacuum infiltration enhances the *Agrobacterium*-mediated genetic transformation in Indian soybean cultivars. *Acta Physiol. Plant* 35, 41–54. doi: 10.1007/s11738-012-1046-3
- Marinho, J. P., Pagliarini, R. F., Molinari, M. D. C., Marcolino-Gomes, J., Caranhoto, A. L. H., Marin, S. R. R., et al. (2022). Overexpression of full-length and partial *DREB2A* enhances soybean drought tolerance. *Agronomy Sci. Biotech.* 8, 1–21. doi: 10.33158/ASB.r141.v8.2022
- Martinell, B. J., Julson, L. A., Hinchee, M. A. W., Connor-Ward, D., McCabe, D., and Emler, C. (1999). *Efficiency Soybean Transformation Protocol. Patent Version No.: US5914451*. St. Louis, Mo: U.S. Patent and Trademark Office, Monsanto Company.
- Martinell, B. J., Julson, L. S., Emler, C. A., Huang, Y., McCabe, D. E., and Williams, E. J. (2002). *Soybean Agrobacterium Transformation Method. Patent Version No.: US6384301*. St. Louis, MO: U.S. Patent and Trademark Office, Monsanto Company.
- Mccabe, D. E., Swain, W. F., Martinell, B. J., and Christou, P. (1988). Stable transformation of soybean (*Glycine Max*) by particle acceleration. *Nat. Biotech.* 6, 923–926. doi: 10.1038/nbt0888-923
- Mehrizadeh, V., Dorani, E., Mohammadi, S. A., and Ghareyazie, B. (2021). Expression of recombinant human IFN- γ protein in soybean (*Glycine max* L.). *Plant Cell Tiss. Org.* 146, 127–136. doi: 10.1007/s11240-021-02052-z
- Meurer, C. A., Dinkins, R. D., and Collins, G. B. (1998). Factors affecting soybean cotyledonary node transformation. *Plant Cell Rep.* 18, 180–186. doi: 10.1007/s002990050553
- Michno, J. M., Wang, X., Liu, J., Curtin, S. J., Kono, T. J., and Stupar, R. M. (2015). CRISPR/Cas mutagenesis of soybean and *Medicago truncatula* using a new web-tool and a modified Cas9 enzyme. *GM Crops Food* 6, 243–252. doi: 10.1080/21645698.2015.1106063
- Mittler, R., Herr, E. H., Orvar, B. L., van Camp, W., Willekens, H., Inzé, D., et al. (1999). Transgenic tobacco plants with reduced capability to detoxify reactive oxygen intermediates are hyperresponsive to pathogen infection. *Proc. Natl. Acad. Sci. U. S. A.* 96, 14165–14170. doi: 10.1073/pnas.96.24.14165
- Moghaieb, R. E. A., Khashaba, E. H., and Abdel Elazim, A. M. (2019). The toxicity of Cry1A5 transgenic soybean plants against *Spodoptera littoralis*. *J. Plant Prot. Res.* 59, 185–191. doi: 10.24425/jppr.2019.129286
- Molinari, M. D. C., Fuganti-Pagliarini, R., Marin, S. R. R., Ferreira, L. C., Barbosa, D. A., Marcolino-Gomes, J., et al. (2020). Overexpression of *AtNCED3* gene improved drought tolerance in soybean in greenhouse and field conditions. *Genet. Mol. Biol.* 43:e20190292. doi: 10.1590/1678-4685-gmb-2019-0292
- Morgante, M., Jurman, I., Shi, L., Zhu, T., Keim, P., and Rafalski, J. A. (1997). The STR120 satellite DNA of soybean: organization, evolution and chromosomal specificity. *Chromosome Res.* 5, 363–373. doi: 10.1023/a:1018492208247
- Mu, R., Lyu, X., Ji, R., Liu, J., Zhao, T., Li, H., et al. (2022). GmBICs modulate low blue light-induced stem elongation in soybean. *Front. Plant Sci.* 13:803122. doi: 10.3389/fpls.2022.803122
- Murashige, T., and Skoog, F. (1962). A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol. Plant.* 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x
- Nadeem, M., Chen, A., Hong, H., Li, D., Li, J., Zhao, D., et al. (2021). GmMsl encodes a kinesin-like protein essential for male fertility in soybean (*Glycine max* L.). *J. Integr. Plant Biol.* 63, 1054–1064. doi: 10.1111/jipb.13110
- Nagy, E. D., Stevens, J. L., Yu, N., Hubmeier, C. S., LaFaver, N., Gillespie, M., et al. (2021). Novel disease resistance gene paralogs created by CRISPR/Cas9 in soy. *Plant Cell Rep.* 40, 1047–1058. doi: 10.1007/s00299-021-02678-5
- Nandula, V. K. (2019). Herbicide resistance traits in maize and soybean: current status and future outlook. *Plants* 8:377. doi: 10.3390/plants8090337
- Ng, M.-S., Ku, Y.-S., Yung, W.-S., Cheng, S.-S., Man, C.-K., Yang, L., et al. (2021). MATE-type proteins are responsible for isoflavone transportation and accumulation in soybean seeds. *Int. J. Mol. Sci.* 22:12017. doi: 10.3390/ijms222112017
- Ngaki, M. N., Sahoo, D. K., Wang, B., and Bhattacharyya, M. K. (2021). Overexpression of a plasma membrane protein generated broad-spectrum immunity in soybean. *Plant Biotechnol. J.* 19, 502–516. doi: 10.1111/pbi.13479
- Nguyen, C., Paddock, K., Zhang, Z., and Stacey, M. (2021). GmKIX8-1 regulates organ size in soybean and is the causative gene for the major seed weight QTL *qSw17-1*. *New Phytol.* 229, 920–934. doi: 10.1111/nph.16928
- Nguyen, H. Q., Le, T. H. T., Nguyen, T. N. L., Nguyen, T. G., Sy, D. T., Tu, Q. T., et al. (2020). Overexpressing *GmCH11A* increases the isoflavone content of transgenic soybean (*Glycine max* (L.) Merr.) seeds. *In Vitro Cell. Dev. Pl.* 56, 842–850. doi: 10.1007/s11627-020-10076-x
- Nguyen, N. T., Vu, H. T., Nguyen, T. T., Nguyen, L.-A. T., Nguyen, M.-C. D., Hoang, K. L., et al. (2019). Co-expression of *Arabidopsis AtAVP1* and *AtNHX1* to improve salt tolerance in soybean. *Crop Sci.* 59, 1133–1143. doi: 10.2135/cropsci2018.10.0640
- Nguyen, Q. H., Vu, L. T. K., Nguyen, L. T. N., Pham, N. T. T., Nguyen, Y. T. H., Le, S. V., et al. (2019). Overexpression of the *GmDREB6* gene enhances proline accumulation and salt tolerance in genetically modified soybean plants. *Sci. Rep.* 9:19663. doi: 10.1038/s41598-019-55895-0
- Nicolas, J. J., Richard-Forget, F. C., Goupy, P. M., Amiot, M. J., and Aubert, S. Y. (1994). Enzymatic browning reactions in apple and apple products. *Crit. Rev. Food Sci.* 34, 109–157. doi: 10.1080/10408399409527653
- Niu, L., Yang, J., Zhang, J., He, H., Xing, G., Zhao, Q., et al. (2019). Introduction of the harpin_{Xoc}-encoding gene *hrf2* in soybean enhances resistance against the oomycete pathogen *Phytophthora sojae*. *Transgenic Res.* 28, 257–266. doi: 10.1007/s11248-019-00119-4
- O’Conner, S., Zheng, W., Qi, M., Kandel, Y., Fuller, R., Whitham, S. A., et al. (2021). GmNF-YC4-2 increases protein, exhibits broad disease resistance and expedites maturity in soybean. *Int. J. Mol. Sci.* 22:3586. doi: 10.3390/ijms22073586
- Olhoft, P. M., Bernal, L. M., Grist, L. B., Hill, D. S., Mankin, S. L., Shen, Y., et al. (2007). A novel *Agrobacterium rhizogenes*-mediated transformation method of soybean [*Glycine max* (L.) Merrill] using primary-node explants from seedlings. *In Vitro Cell. Dev. Pl.* 43, 536–549. doi: 10.1007/s11627-007-9050-9
- Olhoft, P. M., Flagel, L. E., Donovan, C. M., and Somers, D. A. (2003). Efficient soybean transformation using hygromycin B selection in the cotyledonary-node method. *Planta* 216, 723–735. doi: 10.1007/s00425-002-0922-2
- Olhoft, P. M., Flagel, L. E., and Somers, D. A. (2004). T-DNA locus structure in a large population of soybean plants transformed using the *Agrobacterium*-mediated cotyledonary-node method. *Plant Biotechnol. J.* 2, 289–300. doi: 10.1111/j.1467-7652.2004.00070.x
- Olhoft, P. O., Lin, K. L., Galbraith, J. G., Nielsen, N. N., and Somers, D. S. (2001). The role of thiol compounds in increasing *Agrobacterium*-mediated transformation of soybean cotyledonary-node cells. *Plant Cell Rep.* 20, 731–737. doi: 10.1007/s002990100388
- Olhoft, P. O., and Somers, D. S. (2001). L-Cysteine increases *Agrobacterium*-mediated T-DNA delivery into soybean cotyledonary-node cells. *Plant Cell Rep.* 20, 706–711. doi: 10.1007/s002990100379
- Opabode, J. T. (2006). *Agrobacterium*-mediated transformation of plants: emerging factors that influence efficiency. *Biotech. Mol. Bio. Rev.* 1, 12–20.
- Paes de Melo, B., Lourenço-Tessutti, I., Morgante, C., Santos, N., Pinheiro, L., Lins, C., et al. (2020). Soybean embryonic axis transformation: combining biolistic and *Agrobacterium*-mediated protocols to overcome typical complications of *in vitro* plant regeneration. *Front. Plant Sci.* 11:1228. doi: 10.3389/fpls.2020.01228
- Pagano, M. C., and Miransari, M. (2016). “The importance of soybean production worldwide,” in *Abiotic and Biotic Stresses in Soybean Production*, ed. M. Miransari (San Diego: Academic Press), 1–26.
- Pareddy, D., Chennareddy, S., Anthony, G., Sardesai, N., Mall, T., Minnicks, T., et al. (2020). Improved soybean transformation for efficient and high throughput transgenic production. *Transgenic Res.* 29, 267–281. doi: 10.1007/s11248-020-00198-8
- Park, J. S., Kim, H. J., Cho, H. S., Jung, H. W., Cha, J. Y., Yun, D. J., et al. (2019). Overexpression of *AtYUCCA6* in soybean crop results in reduced ROS production and increased drought tolerance. *Plant Biotechnol. Rep.* 13, 161–168. doi: 10.1007/s11816-019-00527-2

- Parrott, W. A., Williams, E. G., Hildebrand, D. F., and Collins, G. B. (1989). Effect of genotype on somatic embryogenesis from immature cotyledons of soybean. *Plant Cell Tiss. Org.* 16, 15–21. doi: 10.1007/BF00044068
- Patil, G. B., Stupar, R. M., and Zhang, F. (2022). "Protoplast Isolation, transfection, and gene editing for soybean (*Glycine max*)," in *Proc. Natl. Acad. Sci.*, eds K. Wang and F. Zhang (New York, NY: Springer US), 173–186. doi: 10.1007/978-1-0716-2164-6_13
- Paz, M. M., Martinez, J. C., Kalvig, A. B., Fonger, T. M., and Wang, K. (2006). Improved cotyledonary node method using an alternative explant derived from mature seed for efficient *Agrobacterium*-mediated soybean transformation. *Plant Cell Rep.* 25, 206–213. doi: 10.1007/s00299-005-0048-7
- Paz, M. M., Shou, H., Guo, Z., Zhang, Z., Banerjee, A. K., and Wang, K. (2004). Assessment of conditions affecting *Agrobacterium*-mediated soybean transformation using the cotyledonary node explant. *Euphytica* 136, 167–179. doi: 10.1023/B:EUPH.0000030669.75809.dc
- Pham, T. T. N., Nguyen, H. Q., Nguyen, T. N. L., Dao, X. T., Sy, D. T., Le, V. S., et al. (2020). Overexpression of the *GmDREB2* gene increases proline accumulation and tolerance to drought stress in soybean plant. *Australian J. Crop Sci.* 14, 495–503. doi: 10.21475/ajcs.20.14.03.p2173
- Pottinger, S. E., Bak, A., Margets, A., Helm, M., Tang, L., Casteel, C., et al. (2020). Optimizing the PBS1 decoy system to confer resistance to potyvirus infection in *Arabidopsis* and soybean. *Mol. Plant Microbe Interact.* 33, 932–944. doi: 10.1094/MPMI-07-19-0190-R
- Punjabi, M., Bharadvaja, N., Jolly, M., Dahuja, A., and Sachdev, A. (2018). Development and evaluation of low phytic acid soybean by siRNA triggered seed specific silencing of inositol polyphosphate 6-/3-/5-kinase gene. *Front. Plant Sci.* 9:804. doi: 10.3389/fpls.2018.00804
- Qin, D., Liu, X. Y., Miceli, C., Zhang, Q., and Wang, P. W. (2019). Soybean plants expressing the *Bacillus thuringiensis* cry8-like gene show resistance to *Holotrichia parallela*. *BMC Biotechnol.* 19:66. doi: 10.1186/s12896-019-0563-1
- Qu, S., Jiao, Y., Abraham, L., and Wang, P. (2021). Correlation analysis of new soybean [*Glycine max* (L.) Merr] gene *Gm15G117700* with oleic acid. *Phyton* 90, 1177–1192. doi: 10.32604/phyton.2021.015206
- Queiroz, L. N., Maldaner, F. R., Mendes, E. A., Sousa, A. R., D'Allastta, R. C., Mendonca, G., et al. (2019). Evaluation of lettuce chloroplast and soybean cotyledon as platforms for production of functional bone morphogenetic protein 2. *Transgenic Res.* 28, 213–224. doi: 10.1007/s11248-019-00116-7
- Rakesh, B., Sudheer, W. N., and Nagella, P. (2021). Role of polyamines in plant tissue culture: an overview. *Plant Cell Tiss. Org.* 145, 487–506. doi: 10.1007/s11240-021-02029-y
- Ramesh, S. V., Shivakumar, M., Praveen, S., Chouhan, B. S., and Chand, S. (2019). Expression of short hairpin RNA (shRNA) targeting AC2 gene of Mungbean yellow mosaic India virus (MYMIV) reduces the viral titre in soybean. *3 Biotech* 9:334. doi: 10.1007/s13205-019-1865-7
- Ran, Y., Liang, Z., and Gao, C. (2017). Current and future editing reagent delivery systems for plant genome editing. *Sci. China Life Sci.* 60, 490–495. doi: 10.1007/s11247-017-9022-1
- Raza, G., Singh, M., and Bhalla, P. (2020). Somatic embryogenesis and plant regeneration from commercial soybean cultivars. *Plants* 9:38. doi: 10.3390/plants9010038
- Rech, E. L., Vianna, G. R., and Aragao, F. J. (2008). High-efficiency transformation by biolistics of soybean, common bean and cotton transgenic plants. *Nat. Protoc.* 3, 410–418. doi: 10.1038/nprot.2008.9
- Rehman, N. U., Abbas, F., Imran, M., Alam, I., Imran, M., Ullah, I., et al. (2022). Genome wide analysis of *DWARF27* genes in soybean and functional characterization of *GmD27c* reveals eminent role of strigolactones in rhizobia interaction and nodulation in *Glycine max*. *Mol. Bio. Rep.* [Epub ahead of print]. doi: 10.1007/s11033-022-07127-4
- Ribichich, K. F., Chiozza, M., Avalos-Britez, S., Cabello, J. V., Arce, A. L., Watson, G., et al. (2020). Successful field performance in warm and dry environments of soybean expressing the sunflower transcription factor HB4. *J. Exp. Bot.* 71, 3142–3156. doi: 10.1093/jxb/eraa064
- Roustan, J. P., Latche, A., and Fallot, J. (1990). Inhibition of ethylene production and stimulation of carrot somatic embryogenesis by salicylic acid. *Biol. Plant.* 32, 273–276. doi: 10.1007/BF02886947
- Sander, J. D., Dahlborg, E. J., Goodwin, M. J., Cade, L., Zhang, F., Cifuentes, D., et al. (2011). Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods* 8, 67–69. doi: 10.1038/nmeth.1542
- Sato, H., Yamada, T., Kita, Y., Ishimoto, M., and Kitamura, K. (2007). Production of transgenic plants and their early seed set in Japanese soybean variety, Kariyutaka. *Plant Biotechnol.* 24, 533–536. doi: 10.5511/plantbiotechnology.24.533
- Sato, S., Newell, C., Kolacz, K., Tredo, L., Finer, J., and Hinchey, M. (1993). Stable transformation via particle bombardment in two different soybean regeneration systems. *Plant Cell Rep.* 12, 408–413. doi: 10.1007/BF00234702
- 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
- Shanmugam, S., Zhao, S., Nandy, S., Srivastava, V., and Khodakovskaya, M. (2020). Modification of soybean growth and abiotic stress tolerance by expression of truncated ERECTA protein from *Arabidopsis thaliana*. *PLoS One* 15:e0233383. doi: 10.1371/journal.pone.0233383
- Shen, Y., Liu, J., Geng, H., Zhang, J., Liu, Y., Zhang, H., et al. (2018). *De novo* assembly of a Chinese soybean genome. *Sci. China Life Sci.* 61, 871–884. doi: 10.1007/s11427-018-9360-0
- Shi, W. Y., Du, Y. T., Ma, J., Min, D. H., Jin, L. G., Chen, J., et al. (2018). The WRKY transcription factor GmWRKY12 confers drought and salt tolerance in soybean. *Int. J. Mol. Sci.* 19:4087. doi: 10.3390/ijms19124087
- Shimomura, M., Kanamori, H., Komatsu, S., Namiki, N., Mukai, Y., Kurita, K., et al. (2015). The *Glycine max* cv. enrei genome for improvement of Japanese soybean cultivars. *Int. J. Genomics* 2015: 358127. doi: 10.1155/2015/358127
- Song, J., Toth, K., Montes-Luz, B., and Stacey, G. (2021). Soybean hairy root transformation: a rapid and highly efficient method. *Curr. Protoc.* 1:e195. doi: 10.1002/cpz1.195
- Song, L., Valliyodan, B., Prince, S., Wan, J., and Nguyen, H. T. (2018). Characterization of the *XTH* gene family: new insight to the roles in soybean flooding tolerance. *Int. J. Mol. Sci.* 19:2705. doi: 10.3390/ijms19092705
- Song, Z. Y., Tian, J. I., Fu, W. Z., Li, L., Lu, L. H., Zhou, L., et al. (2013). Screening Chinese soybean genotypes for *Agrobacterium*-mediated genetic transformation suitability. *J. Zhejiang Univ. Sci. B* 14, 289–298. doi: 10.1631/jzus.B1200278
- Soto, N., Delgado, C., Hernández, Y., Rosabal, Y., Ferreira, A., Pujol, M., et al. (2017). Efficient particle bombardment-mediated transformation of Cuban soybean (INCA-Soy-36) using glyphosate as a selective agent. *Plant Cell Tiss. Org.* 128, 187–196. doi: 10.1007/s11240-016-1099-x
- Soto, N., Hernandez, Y., Delgado, C., Rosabal, Y., Ortiz, R., Valencia, L., et al. (2020). Field resistance to *Phakopsora pachyrhizi* and *Colletotrichum truncatum* of transgenic soybean expressing the *NmDef02* plant defensin gene. *Front. Plant Sci.* 11:562. doi: 10.3389/fpls.2020.00562
- Stacey, G., Vodkin, L., Parrott, W. A., and Shoemaker, R. C. (2004). National science foundation-sponsored workshop report. draft plan for soybean genomics. *Plant Physiol.* 135, 59–70. doi: 10.1104/pp.103.037903
- Subramanian, S., Hu, X., Lu, G., Odell, J. T., and Yu, O. (2004). The promoters of two isoflavone synthase genes respond differentially to nodulation and defense signals in transgenic soybean roots. *Plant Mol. Biol.* 54, 623–639. doi: 10.1023/B:PLAN.0000040814.28507.35
- Sugano, S., Hirose, A., Kanazashi, Y., Adachi, K., Hibara, M., Itoh, T., et al. (2020). Simultaneous induction of mutant alleles of two allergenic genes in soybean by using site-directed mutagenesis. *BMC Plant Biol.* 20:513. doi: 10.1186/s12870-020-02708-6
- Sultana, M. S., Frazier, T. P., Millwood, R. J., Lenaghan, S. C., and Stewart, C. N. (2019). Development and validation of a novel and robust cell culture system in soybean (*Glycine max* (L.) Merr.) for promoter screening. *Plant Cell Rep.* 38, 1329–1345. doi: 10.1007/s00299-019-02455-5
- Sun, T., Ma, N., Wang, C., Fan, H., Wang, M., Zhang, J., et al. (2021). A golgi-localized sodium/hydrogen exchanger positively regulates salt tolerance by maintaining higher K^+/Na^+ ratio in soybean. *Front. Plant Sci.* 12:638340. doi: 10.3389/fpls.2021.638340
- Sun, X., Cai, X., Yin, K., Gu, L., Shen, Y., Hu, B., et al. (2021). Wild soybean SNARE proteins BET1s mediate the subcellular localization of the cytoplasmic receptor-like kinases CRCK1s to modulate salt stress responses. *Plant J.* 105, 771–785. doi: 10.1111/tpj.15072
- Sun, X., Hu, Z., Chen, R., Jiang, Q., Song, G., Zhang, H., et al. (2015). Targeted mutagenesis in soybean using the CRISPR-Cas9 system. *Sci. Rep.* 5:10342. doi: 10.1038/srep10342

- Sun, Z., Su, C., Yun, J., Jiang, Q., Wang, L., Wang, Y., et al. (2019). Genetic improvement of the shoot architecture and yield in soya bean plants via the manipulation of GmMIR156b. *Plant Biotechnol. J.* 17, 50–62. doi: 10.1111/pbi.12946
- Testroet, A., Lee, K., Luth, D., and Wang, K. (2017). Comparison of transformation frequency using the bar gene regulated by the CaMV 35S or NOS promoter in *Agrobacterium*-mediated soybean (*Glycine max* L.) transformation. *In Vitro Cell. Dev. Pl.* 53, 188–199. doi: 10.1007/s11627-017-9810-0
- Tian, B., Li, J., Vodkin, L. O., Todd, T. C., Finer, J. J., and Trick, H. N. (2019). Host-derived gene silencing of parasite fitness genes improves resistance to soybean cyst nematodes in stable transgenic soybean. *Theor. Appl. Genet.* 132, 2651–2662. doi: 10.1007/s00122-019-03379-0
- Torkamaneh, D., Laroche, J., Valliyodan, B., O'Donoghue, L., Cober, E., Rajcan, I., et al. (2019). Soybean haplotype map (GmHapMap): a universal resource for soybean translational and functional genomics. *BioRxiv* [preprint] doi: 10.1101/534578
- Trick, H. N., and Finer, J. J. (1998). Sonication-assisted *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merrill] embryogenic suspension culture tissue. *Plant Cell Rep.* 17, 482–488. doi: 10.1007/s002990050429
- Umezawa, T., Sakurai, T., Totoki, Y., Toyoda, A., Seki, M., Ishiwata, A., et al. (2008). Sequencing and analysis of approximately 40,000 soybean cDNA clones from a full-length-enriched cDNA library. *DNA Res.* 15, 333–346. doi: 10.1093/dnares/dsn024
- Veerappa, R., Slocum, R. D., Siegenthaler, A., Wang, J., Clark, G., and Roux, S. J. (2019). Ectopic expression of a pea apyrase enhances root system architecture and drought survival in *Arabidopsis* and soybean. *Plant Cell Environ.* 42, 337–353. doi: 10.1111/pce.13425
- Verma, N., Tiwari, S., Singh, V. P., and Prasad, S. M. (2020). Nitric oxide in plants: an ancient molecule with new tasks. *Plant Growth Regul.* 90, 1–13. doi: 10.1007/s10725-019-00543-w
- Vianna, G. R., Aragao, F. J., and Rech, E. L. (2011). A minimal DNA cassette as a vector for genetic transformation of soybean (*Glycine max*). *Genet. Mol. Biol.* 10, 382–390. doi: 10.4238/vol10-1gm1058
- Voorra, V., Larrea, C., and Bermudez, S. (2020). “Global market report: soybeans,” in *International Institute for Sustainable Development*, ed. S. Baliño (Winnipeg, Manitoba: International Institute for Sustainable Development).
- Wang, B., Sumit, R., Sahu, B. B., Ngaki, M. N., Srivastava, S. K., Yang, Y., et al. (2018a). *Arabidopsis* novel glycine-rich plasma membrane PSS1 protein enhances disease resistance in transgenic soybean plants. *Plant Physiol.* 176, 865–878. doi: 10.1104/pp.16.01982
- Wang, L. S., Chen, Q. S., Xin, D. W., Qi, Z. M., Zhang, C., Li, S. N., et al. (2018b). Overexpression of *GmBIN2*, a soybean glycogen synthase kinase 3 gene, enhances tolerance to salt and drought in transgenic *Arabidopsis* and soybean hairy roots. *J. Integr. Agr.* 17, 1959–1971. doi: 10.1016/S2095-3119(17)61863-X
- Wang, Y., Jiang, L., Chen, J., Tao, L., An, Y., Cai, H., et al. (2018c). Overexpression of the alfalfa WRKY11 gene enhances salt tolerance in soybean. *PLoS One* 13:e0192382. doi: 10.1371/journal.pone.0192382
- Wang, G., and Xu, Y. (2008). Hypocotyl-based *Agrobacterium*-mediated transformation of soybean (*Glycine max*) and application for RNA interference. *Plant Cell Rep.* 27, 1177–1184. doi: 10.1007/s00299-008-0535-8
- Wang, Y., Yang, Z., Kong, Y., Li, X., Li, W., Du, H., et al. (2020d). *GmPAP12* is required for nodule development and nitrogen fixation under phosphorus starvation in soybean. *Front. Plant Sci.* 11:450. doi: 10.3389/fpls.2020.00450
- Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y.-C., et al. (2020c). Simultaneous changes in seed size, oil content, and protein content driven by selection of SWEET homologues during soybean domestication. *Nat. Sci. Rev.* 7, 1776–1786. doi: 10.1093/nsr/nwaa110
- Wang, L., Sun, S., Wu, T., Liu, L., Sun, X., Cai, Y., et al. (2020b). Natural variation and CRISPR/Cas9-mediated mutation in *GmPRR37* affect photoperiodic flowering and contribute to regional adaptation of soybean. *Plant Biotechnol. J.* 18, 1869–1881. doi: 10.1111/pbi.13346
- Wang, J., Kuang, H., Zhang, Z., Yang, Y., Yan, L., Zhang, M., et al. (2020a). Generation of seed lipooxygenase-free soybean using CRISPR-Cas9. *Crop J.* 8, 432–439. doi: 10.1016/j.cj.2019.08.008
- Wang, Y., Yuan, L., Su, T., Wang, Q., Gao, Y., Zhang, S., et al. (2020e). Light- and temperature-entrainable circadian clock in soybean development. *Plant Cell Environ.* 43, 637–648. doi: 10.1111/pce.13678
- Wang, X., Li, M.-W., Wong, F.-L., Luk, C.-Y., Chung, C. Y.-L., Yung, W.-S., et al. (2021d). Increased copy number of gibberellin 2-oxidase 8 genes reduced trailing growth and shoot length during soybean domestication. *Plant J.* 107, 1739–1755. doi: 10.1111/tpj.15414
- Wang, R., Deng, M., Yang, C., Yu, Q., Zhang, L., Zhu, Q., et al. (2021b). A Qa-SNARE complex contributes to soybean cyst nematode resistance via regulation of mitochondria-mediated cell death. *J. Exp. Bot.* 72, 7145–7162. doi: 10.1093/jxb/erab301
- Wang, W., Chen, L., Fengler, K., Bolar, J., Llaca, V., Wang, X., et al. (2021c). A giant NLR gene confers broad-spectrum resistance to *Phytophthora sojae* in soybean. *Nat. Commun.* 12:6263. doi: 10.1038/s41467-021-26554-8
- Wang, X., Zhao, J., Fang, Q., Chang, X., Sun, M., Li, W., et al. (2021e). *GmAKT1* is involved in K⁺ uptake and Na⁺/K⁺ homeostasis in *Arabidopsis* and soybean plants. *Plant Sci.* 304, 110736. doi: 10.1016/j.plantsci.2020.110736
- Wang, K., Bu, T., Cheng, Q., Dong, L., Su, T., Chen, Z., et al. (2021a). Two homologous LHY pairs negatively control soybean drought tolerance by repressing the abscisic acid responses. *New Phytol.* 229, 2660–2675. doi: 10.1111/nph.17019
- Wang, Y., Yang, W., Zuo, Y., Zhu, L., Hastwell, A. H., Chen, L., et al. (2019c). GmYUC2a mediates auxin biosynthesis during root development and nodulation in soybean. *J. Exp. Bot.* 70, 3165–3176. doi: 10.1093/jxb/erz144
- Wang, L., Wang, H., He, S., Meng, F., Zhang, C., Fan, S., et al. (2019a). GmSnRK1.1, a sucrose non-fermenting-1(SNF1)-related protein kinase, promotes soybean resistance to *Phytophthora sojae*. *Front. Plant Sci.* 10:996. doi: 10.3389/fpls.2019.00996
- Wang, S., Yokosho, K., Guo, R., Whelan, J., Ruan, Y. L., Ma, J. F., et al. (2019b). The soybean sugar transporter GmSWEET15 mediates sucrose export from endosperm to early embryo. *Plant Physiol.* 180, 2133–2141. doi: 10.1104/pp.19.00641
- Wang, P., and Zhao, F.-J. (2019). Engineering crops without genome integration using nanotechnology. *Trends Plant Sci.* 24, 574–577. doi: 10.1016/j.tplants.2019.05.004
- Wang, W., Wang, Z., Hou, W., Chen, L., Jiang, B., Ma, W., et al. (2022). GmNHC5 may promote nodulation via interaction with GmGAI in soybean. *Crop J.* 10, 273–279. doi: 10.1016/j.cj.2021.03.019
- Wang, Z., Wang, Y., Shang, P., Yang, C., Yang, M., Huang, J., et al. (2022). Overexpression of soybean *GmWRI1a* stably increases the seed oil content in soybean. *Int. J. Mol. Sci.* 23:5084. doi: 10.3390/ijms23095084
- Wei, P., Che, B., Shen, L., Cui, Y., Wu, S., Cheng, C., et al. (2019). Identification and functional characterization of the chloride channel gene, *GsCLC-c2* from wild soybean. *BMC Plant Biol.* 19:121. doi: 10.1186/s12870-019-1732-z
- Wei, W., Liang, D. W., Bian, X. H., Shen, M., Xiao, J. H., Zhang, W. K., et al. (2019). *GmWRKY54* improves drought tolerance through activating genes in abscisic acid and Ca²⁺ signaling pathways in transgenic soybean. *Plant J.* 100, 384–398. doi: 10.1111/tpj.14449
- Wu, F., and Hanzawa, Y. (2018). A simple method for isolation of soybean protoplasts and application to transient gene expression analyses. *J. Vis. Exp.* 131:e57258. doi: 10.3791/57258
- Wu, N., Lu, Q., Wang, P., Zhang, Q., Zhang, J., Qu, J., et al. (2020). Construction and analysis of *GmFAD2-1A* and *GmFAD2-2A* soybean fatty acid desaturase mutants based on CRISPR/Cas9 technology. *Int. J. Mol. Sci.* 21:1104. doi: 10.3390/ijms21031104
- Xian, P., Cai, Z., Cheng, Y., Lin, R., Lian, T., Ma, Q., et al. (2020). Wild soybean oxalyl-CoA synthetase degrades oxalate and affects the tolerance to cadmium and aluminum stresses. *Int. J. Mol. Sci.* 21:8869. doi: 10.3390/ijms21228869
- Xiao, P. Y., Liu, Y., and Cao, Y. P. (2019). Overexpression of *G10-EPSPS* in soybean provides high glyphosate tolerance. *J. Integr. Agr.* 18, 1851–1858. doi: 10.1016/S2095-3119(18)62124-0
- Xie, M., Chung, C. Y., Li, M. W., Wong, F. L., Wang, X., Liu, A., et al. (2019). A reference-grade wild soybean genome. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-09142-9
- Xing, X., Du, H., Yang, Z., Li, X., Kong, Y., Li, W., et al. (2022). GmSPX8, a nodule-localized regulator confers nodule development and nitrogen fixation under phosphorus starvation in soybean. *BMC Plant Biol.* 22:161. doi: 10.1186/s12870-022-03556-2
- Xiong, L., Li, C., Li, H., Lyu, X., Zhao, T., Liu, J., et al. (2019). A transient expression system in soybean mesophyll protoplasts reveals the formation of cytoplasmic

- GmCRY1 photobody-like structures. *Sci. China Life Sci.* 62, 1070–1077. doi: 10.1007/s11427-018-9496-5
- Xu, H., Zhang, L., Zhang, K., and Ran, Y. (2020). Progresses, challenges, and prospects of genome editing in soybean (*Glycine max*). *Front. Plant Sci.* 11:571138. doi: 10.3389/fpls.2020.571138
- Xu, M., Li, H., Liu, Z.-N., Wang, X.-H., Xu, P., Dai, S.-J., et al. (2021). The soybean CBL-interacting protein kinase, GmCIPK2, positively regulates drought tolerance and ABA signaling. *Plant Physiol. Biochem.* 167, 980–989. doi: 10.1016/j.plaphy.2021.09.026
- Xu, Y., Yan, F., Zong, Y., Li, J., Gao, H., Liu, Y., et al. (2022). Proteomic and lipidomics analyses of high fatty acid *AhDGAT3* transgenic soybean reveals the key lipase gene associated with the lipid internal mechanism. *Genome* 65, 153–164. doi: 10.1139/gen-2021-0043
- Xue, R. G., Xie, H. F., and Zhang, B. (2006). A multi-needle-assisted transformation of soybean cotyledonary node cells. *Biotechnol. Lett.* 28, 1551–1557. doi: 10.1007/s10529-006-9123-6
- Xun, H., Yang, X., He, H., Wang, M., Guo, P., Wang, Y., et al. (2019). Overexpression of *GmKR3*, a TIR-NBS-LRR type R gene, confers resistance to multiple viruses in soybean. *Plant Mol. Biol.* 99, 95–111. doi: 10.1007/s11103-018-0804-z
- Yamada, T., Takagi, K., and Ishimoto, M. (2012). Recent advances in soybean transformation and their application to molecular breeding and genomic analysis. *Breed. Sci.* 61, 480–494. doi: 10.1270/jsbs.61.480
- Yamada, T., Watanabe, S., Arai, M., Harada, K., and Kitamura, K. (2010). Cotyledonary node pre-wounding with a micro-brush increased frequency of *Agrobacterium*-mediated transformation in soybean. *Plant Biotechnol.* 27, 217–220. doi: 10.5511/plantbiotechnology.27.217
- Yan, B., Srinivasa Reddy, M. S., Collins, G. B., and Dinkins, R. D. (2000). *Agrobacterium tumefaciens*-mediated transformation of soybean [*Glycine max* (L.) Merrill.] using immature zygotic cotyledon explants. *Plant Cell Rep.* 19, 1090–1097. doi: 10.1007/s002990000236
- Yan, Q., Si, J., Cui, X., Peng, H., and Dou, D. (2019). The soybean cinnamate 4-hydroxylase gene *GmCAH1* contributes positively to plant defense via increasing lignin content. *Plant Growth Regul.* 88, 139–149. doi: 10.1007/s10725-019-00494-2
- Yang, S., Feng, Y., Zhao, Y., Bai, J., and Wang, J. (2020b). Overexpression of a *Eutrema salsugineum* phosphate transporter gene *EsPHT1;4* enhances tolerance to low phosphorus stress in soybean. *Biotechnol. Lett.* 42, 2425–2439. doi: 10.1007/s10529-020-02968-0
- Yang, X., Yang, J., Li, H., Niu, L., Xing, G., Zhang, Y., et al. (2020c). Overexpression of the chitinase gene *CmCHI1* from *Coniothyrium minitans* renders enhanced resistance to *Sclerotinia sclerotiorum* in soybean. *Transgenic Res.* 29, 187–198. doi: 10.1007/s11248-020-00190-2
- Yang, Y., Yu, T.-F., Ma, J., Chen, J., Zhou, Y.-B., Chen, M., et al. (2020d). The soybean bZIP transcription factor gene *GmbZIP2* confers drought and salt resistance in transgenic plants. *Int. J. Mol. Sci.* 21:670. doi: 10.3390/ijms21020670
- Yang, C., Huang, Y., Lv, W., Zhang, Y., Bhat, J. A., Kong, J., et al. (2020a). GmNAC8 acts as a positive regulator in soybean drought stress. *Plant Sci.* 293:110442. doi: 10.1016/j.plantsci.2020.110442
- Yang, J., Xing, G., Niu, L., He, H., Guo, D., Du, Q., et al. (2018). Improved oil quality in transgenic soybean seeds by RNAi-mediated knockdown of *GmFAD2-1B*. *Transgenic Res.* 27, 155–166. doi: 10.1007/s11248-018-0063-4
- Yang, X., Niu, L., Zhang, W., Yang, J., Xing, G., He, H., et al. (2018). RNAi-mediated SMV P3 cistron silencing confers significantly enhanced resistance to multiple Potyvirus strains and isolates in transgenic soybean. *Plant Cell Rep.* 37, 103–114. doi: 10.1007/s00299-017-2186-0
- Yang, J., Xing, G., Qian, D. U., Sui, L., Guo, D., Niu, L., et al. (2016). Effects of different soybean genotypes on the transformation efficiency of soybean and analysis of the T-DNA insertions in the soybean genome. *Soybean Sci.* 35, 562–567. doi: 10.11861/j.issn.1000-9841.2016.04.0562
- Yang, X. F., Yu, X. Q., Zhou, Z., Ma, W. J., and Tang, G. X. (2016). A high-efficiency *Agrobacterium tumefaciens* mediated transformation system using cotyledonary node as explants in soybean (*Glycine max* L.). *Acta Physiol. Plant.* 38:60. doi: 10.1007/s11738-016-2081-2
- Yang, S., Yanfeng, H., Cheng, Z., Rice, J., Miao, L., Ma, J., et al. (2019). An efficient *Agrobacterium*-mediated soybean transformation method using green fluorescent protein as a selectable marker. *Plant Signal. Behav.* 14, 1–7. doi: 10.1080/15592324.2019.1612682
- Yang, X., Yang, J., Wang, Y., He, H., Niu, L., Guo, D., et al. (2019). Enhanced resistance to sclerotinia stem rot in transgenic soybean that overexpresses a wheat oxalate oxidase. *Transgenic Res.* 28, 103–114. doi: 10.1007/s11248-018-0106-x
- Yang, X., Li, X., Shan, J., Li, Y., Zhang, Y., Wang, Y., et al. (2021). Overexpression of *GmGAMYB* accelerate the transition to flowering and increases plant height in soybean. *Front. Plant Sci.* 12:667242. doi: 10.3389/fpls.2021.667242
- Yang, Z., Du, H., Xing, X., Li, W., Kong, Y., Li, X., et al. (2022). A small heat shock protein, GmHSP17.9, from nodule confers symbiotic nitrogen fixation and seed yield in soybean. *Plant Biotechnol. J.* 20, 103–115. doi: 10.1111/pbi.13698
- Yao, S. (2001). *Optimization of Agrobacterium-mediated Genetic Transformation of Soybean Using Glufosinate as a Selective Agent*. Ph.D. thesis. Baton Rouge, LA: Louisiana State University. https://digitalcommons.lsu.edu/gradschool_disstheses/329
- Yeom, W., Kim, H., Lee, K.-R., Cho, H., Kim, J.-Y., Jung, H. W., et al. (2020). Increased production of α -linolenic acid in soybean seeds by overexpression of lesquerella *FAD3-1*. *Front. Plant Sci.* 10:1812. doi: 10.3389/fpls.2019.01812
- Yi, J., Derynck, M. R., Li, X., Telmer, P., Marsolais, F., and Dhaubhadel, S. (2010). A single-repeat MYB transcription factor, GmMYB176, regulates *CHS8* gene expression and affects isoflavonoid biosynthesis in soybean. *Plant J.* 62, 1019–1034. doi: 10.1111/j.1365-3113.2010.04214.x
- Yi, X., and Yu, D. (2006). Transformation of multiple soybean cultivars by infecting cotyledonary-node with *Agrobacterium tumefaciens*. *Afr. J. Biotech.* 5, 1989–1993. doi: 10.5897/AJB2006.000-5100
- Yu, G., Zou, J., Wang, J., Zhu, R., Qi, Z., Jiang, H., et al. (2022). A soybean NAC homolog contributes to resistance to *Phytophthora sojae* mediated by dirigent proteins. *Crop J.* 10, 332–341. doi: 10.1016/j.cj.2021.08.009
- Yue, Y., Sun, S., Li, J., Yu, H., Wu, H., Sun, B., et al. (2021). *GmFULa* improves soybean yield by enhancing carbon assimilation without altering flowering time or maturity. *Plant Cell Rep.* 40, 1875–1888. doi: 10.1007/s00299-021-02752-y
- Yun, J., Sun, Z., Jiang, Q., Wang, Y., Wang, C., Luo, Y., et al. (2022). The miR156b-GmSPL9d module modulates nodulation by targeting multiple core nodulation genes in soybean. *New Phytol.* 233, 1881–1899. doi: 10.1111/nph.17899
- Zeng, P., Vadenais, D. A., Zhang, Z., and Polacco, J. C. (2004). Refined glufosinate selection in *Agrobacterium*-mediated transformation of soybean [*Glycine max* (L.) Merrill]. *Plant Cell Rep.* 22, 478–482. doi: 10.1007/s00299-003-0712-8
- Zhang, B., and Xue, R. G. (2019). Refined factors in multi-needle-assisted transformation of soybean. *Asian J. Biotechnol. Genet. Eng.* 2, 1–7.
- Zhang, D., Zhang, H., Hu, Z., Chu, S., Yu, K., Lv, L., et al. (2019a). Artificial selection on *GmOLEO1* contributes to the increase in seed oil during soybean domestication. *PLoS Genet.* 15:e1008267. doi: 10.1371/journal.pgen.1008267
- Zhang, L., Li, T., Wang, Y., Zhang, Y., and Dong, Y. S. (2019b). *FvC5SD* overexpression enhances drought tolerance in soybean by reactive oxygen species scavenging and modulating stress-responsive gene expression. *Plant Cell Rep.* 38, 1039–1051. doi: 10.1007/s00299-019-02424-y
- Zhang, W., Liao, X., Cui, Y., Ma, W., Zhang, X., Du, H., et al. (2019c). A cation diffusion facilitator, GmCDF1, negatively regulates salt tolerance in soybean. *PLoS Genet.* 15:e1007798. doi: 10.1371/journal.pgen.1007798
- Zhang, X. Z., Zheng, W. J., Cao, X. Y., Cui, X. Y., Zhao, S. P., Yu, T. F., et al. (2019d). Genomic analysis of stress associated proteins in soybean and the role of *GmSAP16* in abiotic stress responses in *Arabidopsis* and soybean. *Front. Plant Sci.* 10:1453. doi: 10.3389/fpls.2019.01453
- Zhang, F., Chen, C., Ge, H., Liu, J., Luo, Y., Liu, K., et al. (2014). Efficient soybean regeneration and *Agrobacterium*-mediated transformation using a whole cotyledonary node as an explant. *Biotech. App. Biochem.* 61, 620–625. doi: 10.1002/bab.1207
- Zhang, H., Yang, Y., Sun, C., Liu, X., Lv, L., Hu, Z., et al. (2020). Up-regulating *GmETO1* improves phosphorus uptake and use efficiency by promoting root growth in soybean. *Plant Cell Environ.* 43, 2080–2094. doi: 10.1111/pce.13816
- Zhang, M., Liu, Y., Cai, H., Guo, M., Chai, M., She, Z., et al. (2020). The bZIP transcription factor *GmbZIP15* negatively regulates salt- and drought-stress responses in soybean. *Int. J. Mol. Sci.* 21:7778. doi: 10.3390/ijms21027778
- Zhang, Y., Zhao, Q., Zhang, J., Niu, L., Yang, J., Liu, X., et al. (2022c). Enhanced resistance to soybean cyst nematode in transgenic soybean via host-induced

- silencing of vital *Heterodera glycines* genes. *Transgenic Res.* 31, 239–248. doi: 10.1007/s11248-022-00298-7
- Zhang, H.-Y., Hou, Z.-H., Zhang, Y., Li, Z.-Y., Chen, J., Zhou, Y.-B., et al. (2022a). A soybean EF-Tu family protein GmEF8, an interactor of GmCBL1, enhances drought and heat tolerance in transgenic *Arabidopsis* and soybean. *Int. J. Biol. Macromol.* 205, 462–472. doi: 10.1016/j.ijbiomac.2022.01.165
- Zhang, Y., Guo, W., Chen, L., Shen, X., Yang, H., Fang, Y., et al. (2022b). CRISPR/Cas9-mediated targeted mutagenesis of *GmUGT* enhanced soybean resistance against leaf-chewing insects through flavonoids biosynthesis. *Front. Plant Sci.* 13:802716. doi: 10.3389/fpls.2022.802716
- Zhang, W. J., Dewey, R. E., Boss, W., Phillippy, B. Q., and Qu, R. (2013). Enhanced *Agrobacterium*-mediated transformation efficiencies in monocot cells is associated with attenuated defense responses. *Plant Mol. Biol.* 81, 273–286. doi: 10.1007/s11103-012-9997-8
- Zhang, Y. M., Liu, Z. H., Yang, R. J., Li, G. L., Guo, X. L., Zhang, H. N., et al. (2016). Improvement of soybean transformation via *Agrobacterium tumefaciens* methods involving α -aminooxyacetic acid and sonication treatments enlightened by gene expression profile analysis. *Plant Cell Rep.* 35, 1259–1271. doi: 10.1007/s00299-016-1958-2
- Zhang, Y.-M., Zhang, H.-M., Liu, Z.-H., Guo, X.-L., Li, H.-C., Li, G.-L., et al. (2015). Inhibition of isoflavone biosynthesis enhanced T-DNA delivery in soybean by improving plant-*Agrobacterium tumefaciens* interaction. *Plant Cell Tiss. Org.* 121, 183–193. doi: 10.1007/s11240-014-0693-z
- Zhang, Z., Xing, A., Staswick, P., and Clemente, T. E. (1999). The use of glufosinate as a selective agent in *Agrobacterium*-mediated transformation of soybean. *Plant Cell Tiss. Org.* 56, 37–46. doi: 10.1023/A:1006298622969
- Zhao, C., Pan, X., Yu, Y., Zhu, Y., Kong, F., Sun, X., et al. (2020). Overexpression of a TIFY family gene, *GsJAZ2*, exhibits enhanced tolerance to alkaline stress in soybean. *Mol. Breed.* 40:33. doi: 10.1007/s11032-020-01113-z
- Zhao, Q., Du, Y., Wang, H., Rogers, H. J., Yu, C., Liu, W., et al. (2019b). 5-Azacytidine promotes shoot regeneration during *Agrobacterium*-mediated soybean transformation. *Plant Physiol. Biochem.* 141, 40–50. doi: 10.1016/j.plaphy.2019.05.014
- Zhao, M. J., Yin, L. J., Ma, J., Zheng, J. C., Wang, Y. X., Lan, J. H., et al. (2019a). The roles of *GmERF135* in improving salt tolerance and decreasing ABA sensitivity in soybean. *Front. Plant Sci.* 10:940. doi: 10.3389/fpls.2019.00940
- Zhao, Y., Zhu, L., Lin, C., Shen, Z., and Xu, C. (2019c). Transgenic soybean expressing a thermostable phytase as substitution for feed additive phytase. *Sci. Rep.* 9:14390. doi: 10.1038/s41598-019-51033-y
- Zhao, X., Jing, Y., Luo, Z., Gao, S., Teng, W., Zhan, Y., et al. (2021). *GmST1*, which encodes a sulfotransferase, confers resistance to soybean mosaic virus strains G2 and G3. *Plant Cell Environ.* 44, 2777–2792. doi: 10.1111/pce.14066
- Zheng, N., Li, T., Dittman, J., Su, J., Li, R., Gassmann, W., et al. (2020). CRISPR/Cas9-based gene editing using egg cell-specific promoters in *Arabidopsis* and soybean. *Front. Plant Sci.* 11:800. doi: 10.3389/fpls.2020.00800
- Zheng, Q., and Perry, S. E. (2014). Alterations in the transcriptome of soybean in response to enhanced somatic embryogenesis promoted by orthologs of Agamous-like15 and Agamous-like18. *Plant Physiol.* 164, 1365–1377. doi: 10.1104/pp.113.234062
- Zheng, T., Yu, X., Sun, Y., Zhang, Q., Zhang, X., Tang, M., et al. (2022). Expression of a cytochrome P450 gene from bermuda grass *Cynodon dactylon* in soybean confers tolerance to multiple herbicides. *Plants* 11:949. doi: 10.3390/plants11070949
- Zhou, Y., Huang, J. L., Zhang, X. L., Zhu, L. M., Wang, X. F., Guo, N., et al. (2018). Overexpression of chalcone isomerase (CHI) increases resistance against *Phytophthora sojae* in soybean. *J. Plant Biol.* 61, 309–319. doi: 10.1007/s12374-018-0017-7
- Zhou, Y., Liu, W., Li, X., Sun, D., Xu, K., Feng, C., et al. (2020). Integration of sRNA, degradome, transcriptome analysis and functional investigation reveals gma-miR398c negatively regulates drought tolerance via *GmCSDs* and *GmCCS* in transgenic *Arabidopsis* and soybean. *BMC Plant Biol.* 20:190. doi: 10.1186/s12870-020-02370-y
- Zhu, S., Chen, Z., Xie, B., Guo, Q., Chen, M., Liang, C., et al. (2021). A phosphate starvation responsive malate dehydrogenase, *GmMDH12* mediates malate synthesis and nodule size in soybean (*Glycine max*). *Environ. Exp. Bot.* 189:104560. doi: 10.1016/j.envexpbot.2021.104560

Conflict of Interest: HX and YR were employed by Tianjin Genovo Biotechnology Co., Ltd.

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 Xu, Guo, Qiu and Ran. 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.



Identification of Candidate Genes and Genomic Selection for Seed Protein in Soybean Breeding Pipeline

Jun Qin¹, Fengmin Wang¹, Qingsong Zhao¹, Ainong Shi^{2*}, Tiantian Zhao¹, Qijian Song³, Waltram Ravelombola⁴, Hongzhou An¹, Long Yan¹, Chunyan Yang^{1*} and Mengchen Zhang^{1*}

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Milind B. Ratnaparkhe,
ICAR Indian Institute of Soybean
Research, India
Hengyou Zhang,
Key Laboratory of Soybean Molecular
Design and Breeding, Northeast
Institute of Geography
and Agroecology (CAS), China
Javaid Akhter Bhat,
Nanjing Agricultural University, China

*Correspondence:

Ainong Shi
ashi@uark.edu
Chunyan Yang
chyyang66@163.com
Mengchen Zhang
zhangmengchend@163.com

Specialty section:

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

Received: 24 February 2022

Accepted: 16 May 2022

Published: 16 June 2022

Citation:

Qin J, Wang F, Zhao Q, Shi A,
Zhao T, Song Q, Ravelombola W,
An H, Yan L, Yang C and Zhang M
(2022) Identification of Candidate
Genes and Genomic Selection
for Seed Protein in Soybean Breeding
Pipeline. *Front. Plant Sci.* 13:882732.
doi: 10.3389/fpls.2022.882732

¹ National Soybean Improvement Center Shijiazhuang Sub-Center, North China Key Laboratory of Biology and Genetic Improvement of Soybean, Ministry of Agriculture, Hebei Laboratory of Crop Genetics and Breeding, Cereal & Oil Crop Institute, Hebei Academy of Agricultural and Forestry Sciences, Shijiazhuang, China, ² Department of Horticulture, University of Arkansas, Fayetteville, AR, United States, ³ Soybean Genomics and Improvement Lab, United States Department of Agriculture - Agricultural Research Service (USDA-ARS), Beltsville, MD, United States, ⁴ Department of Soil and Crop Sciences, Texas A&M University, College Station, TX, United States

Soybean is a primary meal protein for human consumption, poultry, and livestock feed. In this study, quantitative trait locus (QTL) controlling protein content was explored via genome-wide association studies (GWAS) and linkage mapping approaches based on 284 soybean accessions and 180 recombinant inbred lines (RILs), respectively, which were evaluated for protein content for 4 years. A total of 22 single nucleotide polymorphisms (SNPs) associated with protein content were detected using mixed linear model (MLM) and general linear model (GLM) methods in Tassel and 5 QTLs using Bayesian interval mapping (IM), single-trait multiple interval mapping (SMIM), single-trait composite interval mapping maximum likelihood estimation (SMLE), and single marker regression (SMR) models in Q-Gene and IciMapping. Major QTLs were detected on chromosomes 6 and 20 in both populations. The new QTL genomic region on chromosome 6 (Chr6_18844283–19315351) included 7 candidate genes and the Hap.X^{AA} at the Chr6_19172961 position was associated with high protein content. Genomic selection (GS) of protein content was performed using Bayesian Lasso (BL) and ridge regression best linear unbiased prediction (rrBLUP) based on all the SNPs and the SNPs significantly associated with protein content resulted from GWAS. The results showed that BL and rrBLUP performed similarly; GS accuracy was dependent on the SNP set and training population size. GS efficiency was higher for the SNPs derived from GWAS than random SNPs and reached a plateau when the number of markers was >2,000. The SNP markers identified in this study and other information were essential in establishing an efficient marker-assisted selection (MAS) and GS pipelines for improving soybean protein content.

Keywords: *Glycine max*, genome-wide association study, genomic selection, genotyping by sequencing, protein content, single nucleotide polymorphism

INTRODUCTION

Soybean [*Glycine max* (L.) Merr.] provides about 60% of the vegetable-derived proteins worldwide and is a primary meal protein for human consumption, poultry, and livestock feed (Wolf, 1970; Patil et al., 2017). Improving protein content is one of the major breeding objectives in breeding programs (Li S. et al., 2019; Stewart-Brown et al., 2019). Traditional soybean breeding methods require phenotyping and multigeneration selection. Although molecular marker-assisted selection (MAS) by tagging the desired genes during breeding selection is an approach to make the selection more efficient (Collard et al., 2005), it is only relatively effective for traits with high heritability and controlled by major genes (Xu and Crouch, 2008; Xu et al., 2012; Patil et al., 2017). Genomic selection (GS) was developed for the selection of traits controlled by multiple genes, but it has not been practically applied due to the large variation of prediction accuracy in different populations and lacking efficient genotyping platforms (Zhang A. et al., 2017; Liu et al., 2018). With the rapid development of genomic tools and DNA sequencing technology, breeders and geneticists are able to explore molecular approaches to increase seed protein genetic gain (Song et al., 2004, 2013; Schmutz et al., 2010; Wang et al., 2020).

Linkage analysis (Hyten et al., 2004; Nichols et al., 2006; Pathan et al., 2013; Teng et al., 2017; Whiting et al., 2020) and genome-wide association study (GWAS) are powerful tools to identify markers associated with seed protein content in soybean (Hwang et al., 2014; Leamy et al., 2017; Lee et al., 2019; Li S. et al., 2019); to date, a total of 262 loci have been reported through linkage analysis and 107 loci have been reported through GWAS (Patil et al., 2017; Gangurde et al., 2020) per SoyBase.¹ These loci were on all the chromosomes, especially chromosome (Chr.) 15 and Chr. 20 (see text footnote 1). Among these, several quantitative trait loci (QTLs), such as *cqPro-20* on Chr. 20 and *cqPro-15* on Chr. 15, were confirmed based on a low error rate (lower than 0.01) and in different populations (Patil et al., 2017). More than 150 candidate genes have been suggested to control seed protein content in soybean (Zhang D. et al., 2017; Zhang J. et al., 2018; Zhang Y. et al., 2018; Li S. et al., 2019; Zhang et al., 2019; Wang et al., 2020). The most described genes affecting seed protein content were sugar efflux transporter SWEET39 (*Glyma15g05470*) and sugar efflux transporter SWEET24 (*Glyma08g19580*) (Wang et al., 2020).

The populations used for mapping protein content in the previous reports included pedigree-based F2 and F4:6 (Csanádi et al., 2001; Chapman et al., 2003), recombinant inbred lines (RILs) population (Qi et al., 2014; Hacisalihoglu et al., 2018), backcross population (Sebolt et al., 2000; Liang et al., 2010), multiline population (Brummer et al., 1997; Wang et al., 2014; Whiting et al., 2020), nested association mapping population (Gangurde et al., 2020), and natural population (Hwang et al., 2014; Bandillo et al., 2015; Li D. et al., 2019). Most studies used a single population, but some studies used two populations for QTL verification (Vaughn et al., 2014; Zhang D. et al., 2017; Zhang et al., 2019); a few studies analyzed

QTL using both the linkage mapping and associate mapping methods (Zhang et al., 2019).

The annual wild soybean (*Glycine soja*) is an important resource to improve soybean (Lam et al., 2010; Yao et al., 2020). Therefore, the objectives of this study were to: (1) identify QTL conferring seed protein content in RILs derived from cultivated and wild soybeans; (2) identify single nucleotide polymorphism (SNP) markers associated with seed protein content in GWAS and candidate genes controlling the trait; and (3) assess the accuracy of GS base on different SNP sets, training population size, and statistical models.

MATERIALS AND METHODS

Plant Materials

Recombinant Inbred Line

A population of 180 F9-derived RILs was developed from a cross of Jidou12 (*Glycine max*) and Ye9 (*Glycine soja*). Jidou12 is a high-yield cultivar with a high protein content that is grown in Shandong Jiaodong Peninsula, Hebei Province, and south-central Shanxi. The seed protein content averaged 46.48% for Jidou12 and 48.78% for Ye9 on a dry weight basis.

Natural Population

A total of 284 soybean genotypes, including 250 accessions selected from germplasm collection by Dr. Lijuan Qiu's laboratory at the Chinese Academy of Agricultural Sciences and 34 cultivars from Hebei Province, were used for the protein association analysis (**Supplementary Table 1**). These genotypes were originally from 10 provinces in China (202, 67.5%) and 6 states in the United States (76, 30.1%), South Korea (3, 1.2%), and Japan (2, 0.8%).

Field Design

Field experiments were conducted at Shijiazhuang (114°83'E, 38°03'N) in Hebei Province in a randomized complete block design with three replications in 2008, 2010, 2019, and 2020. The plot size was 3 m × 6 m with six rows and 50 cm space between rows in all the trials. The planting density was 225,000 plants per ha. Each year, the plots were irrigated once at the seed-filling stage. Plants were harvested after 95% of the leaves were falling off. Ten plants were randomly chosen from the middle of the plot for indoor laboratory seed protein content analysis when 95% of plants in the plot were matured.

Statistical Analysis of Phenotypic Data

Seed protein content was quantified using Fourier transform-near IR spectroscopy (Bruker MPA, Karlsruhe, Germany) at the North China Key Laboratory of Biology and Genetic Improvement of Soybean, Ministry of Agriculture. Under the Quant 2 method of OPUS (<https://www.bruker.com/en/products-and-solutions/infrared-and-raman/opus-spectroscopy-software/downloads.html>) version 5.5 software (Bruker MPA, Karlsruhe, Germany), the samples' protein content data were calculated using the dry basis model (Yan et al., 2008). Each RIL and accession from each replication

¹<https://www.soybase.org/>

of each environment were detected three times using about 100–150 dry seeds and the average was used for statistical analysis. Analysis of variance was performed using JMP® (https://www.jmp.com/en_us/home.html) Genomics 7 (Sall et al., 2017). The least-squares mean (LSM) of the protein content of each soybean genotype from JMP was used as the phenotypic data in the association mapping.

Genotyping by Sequencing and Single Nucleotide Polymorphism Discovery

Genomic DNA was extracted from leaves of soybean plants using the QIAGEN DNeasy Plant Mini Kit (250). DNA was digested using the restriction enzyme *ApeKI* following the genotyping by sequencing (GBS) protocol described by Elshire et al. (2011). The 90 bp pair-end sequencing of accessions was performed using an Illumina HiSeq 2000 machine at the Genetic Research Institute, Chinese Academy of Sciences. GBS data alignment, mapping, and SNP discovery were done using Short oligonucleotide analysis Package (SOAP) family software. An average of 3.26 M short reads for each accession was aligned to soybean whole-genome sequence (Wm82.a2.v1) using SOAPaligner/soap2. SOAPsnp version 1.05 was used for SNP calling (Li et al., 2009; Li, 2011). Approximately, a half-million SNPs were discovered among the 284 soybean germplasm accessions. The SNPs were filtered before genetic diversity and association analyses. Soybean accession with >5% missing SNP and the >2% heterozygous SNP genotypes was eliminated. After the SNP dataset was filtered to remove those SNPs with minor allele frequency (MAF) <2%, missing data >5%, and heterozygous genotype >25%, a total of 10,115 SNPs were used for genetic diversity and association analysis (Supplementary Figure 1).

Genetic Maps

The genetic maps were constructed with JoinMap 4.0 (Van Ooijen, 2006) when the threshold for the logarithm of odds (LOD) was 3.0 based on 180 F9 RILs. QTL analysis of protein content in the RIL population was performed using single-trait Bayesian interval mapping (BIM), single-trait multiple interval mapping (SMIM), single-trait composite interval mapping maximum likelihood estimator (SMLE), single marker regression (SMR) method of Q-gene software (Joehanes and Nelson, 2008) with inclusive composite interval mapping (ICIM, <http://www.isbreeding.net>) (Meng et al., 2015). Variance components, QTL heritability, and QTL effect for seed protein content were estimated by QTLNetwork version 2.1 based on the phenotypic data (Yang et al., 2008). Only the QTL, which was mapped in similar physical locations (<1,500 kb) on the same chromosomes based on the five methods, was defined as a reliable QTL. The selected SNP markers were further tested for their effect by variance analysis using JMP Pro 10 (Sall et al., 2017).

Population Genetic Diversity and Association Analysis

STRUCTURE is a program that uses Bayesian methods to analyze multilocus data in population genetics (Kaeuffer et al., 2007). This study used a hybrid model and an allelic variation

occurrence non-correlative model to examine the population structure of soybean germplasm. The number of the subpopulation (K) was assumed to be between 1 and 12. Each K was run 10 times, the Markov Chain Monte Carlo (MCMC) length of the burn-in period was 20,000, and the number of the MCMC iterations after the burn-in was 50,000. Delta K was used to determine appropriate K-values (Earl and vonHoldt, 2012). Next, CLUMPP was used to integrate the STRUCTURE-generated results with the “repeat 1,000” parameter. In addition, two different association mapping models were used to analyze the association between the molecular markers and traits, the TASSEL general linear model (GLM-Q), and the mixed linear model (MLM) combining kinship with population structure (Q-matrix) (Yu et al., 2006; Bradbury et al., 2007).

Identification of Candidate Genes

Linkage disequilibrium (LD) analysis was performed in the regions with SNP significantly associated with protein content; SNPs with $r^2 > 0.5$ in a 1-Mb window were considered to be in one linkage disequilibrium (LD) block in the heterochromatic regions. Haplotype analysis was conducted on all the SNPs within the LD block containing significant loci. Two databases, namely, the SoyBase (see text footnote 1) and the *Arabidopsis* Information Resource,² were used for gene annotation and preliminary screening of candidate genes were determined by combined bioinformatics and statistics.

Genomic Selection

Ridge regression best linear unbiased prediction (rrBLUP) and Bayesian Lasso Regression (BLR) were used to predict genomic estimated breeding value (GEBV) in GS (Endelman, 2011; Legarra et al., 2011). The packages “rrBLUP” (Endelman, 2011) and “BGLR” (Pérez and de los Campos, 2014) containing the GS models rrBLUP and Bayesian Lasso (BL), respectively, were run in R software.

Prediction accuracy of seed protein was evaluated for different SNP sets, including 22 significant SNPs detected from GWAS, 22 random SNPs, 100 random SNPs, 250 random SNPs, 500 random SNPs, 1,000 random SNPs, 2,000 random SNPs, 5,000 random SNPs, and 10,115 SNPs. The effect of training population size on GS accuracy was investigated by conducting cross-validation at different levels with 100 replications for each cross-validation fold from two to ten.

RESULTS

Seed Protein Content Variations in Two Populations

The seed protein content of the 180 RILs showed a biased normal distribution, seed protein content ranged from 34.69 to 58.71, and the Coefficient of variation (CV) was 23.39% (Supplementary Figure 2A). The seed protein content of the 284 accessions showed a biased normal distribution, seed

²<https://www.arabidopsis.org/>

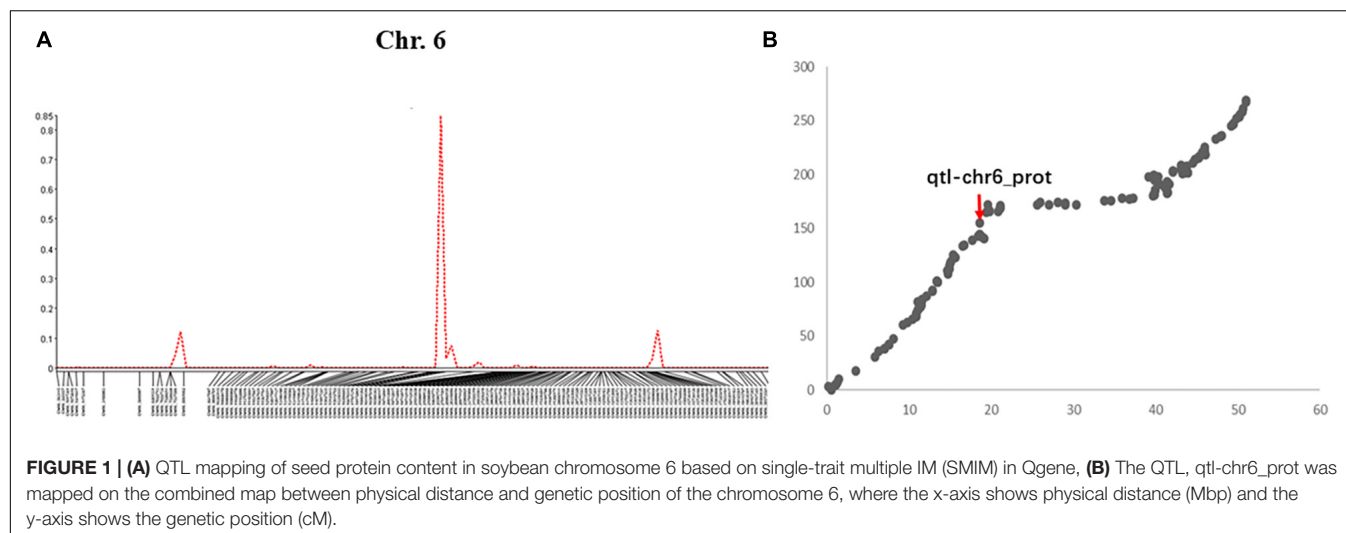
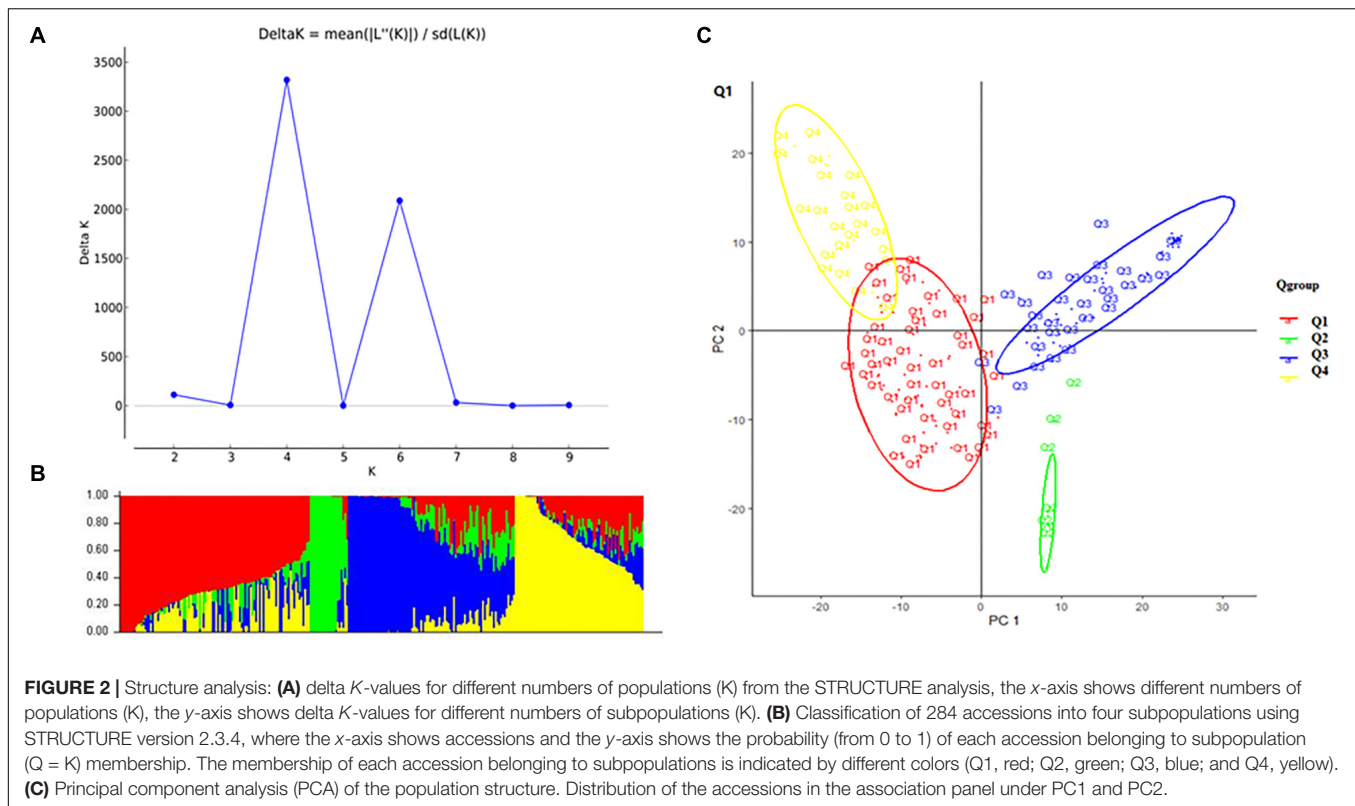


TABLE 1 | Single nucleotide polymorphism (SNP) markers/quantitative trait locus (QTL) detected in recombinant inbred line (RIL) and natural populations.

SNP Markers/QTL Detected in RIL and Natural populations	Population	Model	Confidence interval	Physical position bp	LOD	Posterior (POP)	PVE (%)
qtl-chr6_prot	RIL	Bayesian IM	142	18864382		0.847	
		Single-trait multiple IM (SMIM)	146–152	18580363–18597849	11.46		25.40
		Single-trait CIM MLE (SMLE)	142–152	18580363–18864382	13.1		
		Single marker regression (SMR)	141.9–144.5	18449510–19398117	13.7		29.60
		ICIM	144	18449510–18597849	14.11		22.30
Chr6_18658898	POP	MLM		18658898	19.95		
		GLM		18658898	25.76		
qtl-chr8_prot	RIL	Bayesian IM	56–58	9318625–9502316		0.392–0.49	
		Single-trait multiple IM (SMIM)	42–44	7270752–8285888	7.16		16.70
		Single-trait CIM MLE (SMLE)	42–44	7270752–8285888	7.05		
		Single marker regression	41.6–45.6	7270752–8285888	6.79		16.10
		ICIM	61	9701254–9877332	6.35		9.18
qtl-chr15_prot	RIL	Bayesian IM	12	1890050		0.179	
			32	4708800–4708818		0.759	
			42	5786875		0.119	
		Single-trait multiple IM (SMIM)	20–32	3303648–4708818	3.28		8.00
		Single-trait CIM MLE (SMLE)	18–20	3380704–3303648	4.43		
			30–50	4708800–6651199	4.93		
		Single marker regression	14.2–19.7	2095208–3303648	4.57		11.00
			27.7–31.8	4370908–4708800	3.91		9.50
			45.1–54.5	6037184–7193889	4.43		10.70
		ICIM	20	3303648–3488588	4.72		6.60
qtl-chr17_prot1	RIL	Bayesian IM	100	12398690–12801544		0.941	
		Single-trait multiple IM (SMIM)	100–124	12398690–13632893	4.11		9.80
		Single-trait CIM MLE (SMLE)	104–112	12801549–13813134	4.05		9.90
		Single marker regression	99.5–103.9	12398690–12801549	4.4		
qtl-chr20_prot	RIL	Bayesian IM	112	33202705		0.871	
		Single-trait multiple IM (SMIM)	94	33202705	6.31		14.90
		Single-trait CIM MLE (SMLE)	86–114	26572911–33224754	5.34		
		Single marker regression	93–115.3	26572981–33507017	5.12		12.30
		ICIM	97	26957096–27003724	7.16		10.22
Chr20_34423091	POP	MLM		34423091	7.21		
Chr20_34423091		GLM		34423091	6.55		



protein ranged from 35.65 to 50.99, and the CV was 9.53% (Supplementary Figure 2B).

Genetic Map Construction and Quantitative Trait Locus Mapping in Recombinant Inbred Line Population

The RIL population was genotyped by sequencing. After filtering, a total of 2,498 polymorphic markers SNP were obtained and were mapped to 20 soybean chromosomes, thus the genetic maps were built for the RILs (Supplementary Figure 3A). According to their physical positions in the genome assembly, these markers were basically evenly distributed on 20 chromosomes. The 20 combined maps between physical distance and genetic position showed a good match (Supplementary Figure 3B). Chr. 14 had the least number of markers (68) and Chr. 18 had the largest number of markers (184). A genetic linkage map with a total length of 4,476.2 cm was constructed and the average distance between two adjacent markers was 1.8 cm (Supplementary Figure 3). The average distance between adjacent markers was the smallest on Chr. 20 (1.32 cm) but was the largest on Chr. 9 (2.26 cm).

A total of 5 QTLs on chromosomes 6, 8, 15, 17, and 20 were detected and the LOD value of the markers associated with the QTL ranged from 3.3 to 14.1; the QTL could explain 6.6%–29.6% of the genetic variation (Figure 1A and Supplementary Figure 4). Among these, one QTL with a positive allelic effect was from Jidou12 and 4 QTL with positive alleles were from Ye9 (Table 1). The QTL *qtl-chr6-prot* had the highest LOD and could

explain 22.3–29.6% of genetic variation (Table 1 and Figure 1A). The *qtl-chr6-prot* was in the heterochromatic region (Figure 1B).

Genome-Wide Association Study in Natural Population and Candidate Genes Selection

A total of 10,115 high-quality SNPs were used to perform population structure analysis of the 284 accessions using the STRUCTURE software (Kaeuffer et al., 2007). When *K* = 4, delta *K* was maximal with a relatively stable α value (Figures 2A,B). Cluster I was comprised of 102 accessions, including 77 cultivars, 21 landrace, and 5 exotic accessions; cluster II was comprised of 19 accessions, namely, 18 exotic accessions and 1 cultivar; cluster III was comprised of 93 accessions, namely, 57 exotic accessions, 34 cultivars, and 2 landraces; and cluster IV comprised of 70 accessions, namely, 51 landraces, 16 cultivars, and 3 exotic accessions. Principal component analysis (PCA) also showed the four groups (Figure 2C).

A significant association ($-\log P > 5.35$) with seed protein was observed for 22 SNPs from 22 haplotype blocks in 13 of the 20 chromosomes using GLM and MLM (Table 2). The LOD of the 22 markers ranged from 6.6 to 20.1 in GLM analysis and 6.3 to 26.3 in MLM analysis (Table 2 and Supplementary Figure 5), indicating that these markers were strongly associated with seed protein. Eighteen of these markers were in euchromatic regions and four of these markers were in heterochromatic regions (Table 2).

TABLE 2 | Significant SNPs associated with protein content over 4 years, chromosome (Chr.) and physical position (bp) of the significant SNPs, logarithm of odds (LOD) [$-\log_{10}$ (p -value)] values of generalized linear model (GLM) and mixed liner model (MLM), and allele with positive effect at the SNP locus.

SNP Markers	Chr.	Position	Heterochromatic region	Euchromatic region	SNP Type	Allele with positive effect	LOD of GLM	LOD of MLM	SNP annotation
Chr03_34851073	3	34,851,073		E	A/C	C	12.79	13.69	Glyma.03G133300
Chr03_42692363	3	42,692,363		E	C/T	C	10.22	9.18	Glyma.03G224600
Chr05_40074496	5	40,074,496		E	A/T	T	20.03	25.86	Glyma.05G221300
Chr05_41114434	5	41,114,434		E	C/T	C	13.08	13.12	Upstream_gene_variant MODIFIER Glyma.05G234000
Chr06_14606307	6	14,606,307		E	A/G	G	9.30	8.41	Upstream_gene_variant MODIFIER Glyma.06G173600
Chr06_18658898	6	18,658,898	H		A/G	A	19.95	25.76	Glyma.06G202000
Chr08_10757609	8	10,757,609		E	C/T	C	7.23	6.65	Glyma.08G140700
Chr09_5898756	9	5,898,756		E	A/G	G	8.80	8.45	Glyma.09G062100
Chr09_45699847	9	45,699,847		E	A/G	G	8.55	7.64	Glyma.09G234500
Chr10_2992389	10	2,992,389		E	A/T	A	8.47	7.94	Glyma.10G034400
Chr10_44549078	10	44,549,078		E	A/G	G	9.22	8.80	Glyma.10G213000
Chr12_1536444	12	1,536,444		E	A/G	G	7.48	6.29	Glyma.12G021400
Chr14_2351357	14	2,351,357		E	C/T	C	10.58	9.26	Glyma.14G032300
Chr14_48312781	14	48,312,781		E	C/G	G	20.07	26.26	Upstream_gene_variant MODIFIER Glyma.14G218000
Chr15_13541492	15	13,541,492	H		C/G	C	8.48	7.71	Upstream_gene_variant MODIFIER Glyma.15G160000
Chr17_347445	17	347,445		E	A/T	A	9.38	9.40	Glyma.17G003000
Chr17_32480031	17	32,480,031	H		A/G	G	6.64	6.73	Intergenic_region MODIFIER Glyma.17G203300- Glyma.17G203400
Chr18_7837981	18	7,837,981		E	A/C	C	14.27	13.62	Glyma.18G081200
Chr18_18834295	18	18,834,295		E	A/C	C	8.24	7.25	Intergenic_region MODIFIER Glyma.18G133000- Glyma.18G133100
Chr18_50849168	18	50,849,168		E	A/T	A	11.29	11.66	Upstream_gene_variant MODIFIER Glyma.18G221300
Chr19_12210884	19	12,210,884	H		C/T	T	19.91	25.75	Intergenic_region MODIFIER Glyma.19G060900- Glyma.19G061000
Chr20_34423091	20	34,423,091		E	A/T	T	7.21	6.55	Glyma.20G100900

Two significant SNP loci on Chr. 6 and 20 were detected in linkage analysis and GWAS and the SNP loci detected on Chr. 6 by GWAS were in the QTL intervals obtained by linkage analysis. This SNP region on Chr. 6 had a high Phenotypic variation explained (PVE) (22.3–29.60%) and LOD (6.696–25.762). The region on Chr. 20 was associated with protein content with a PVE of 12.30% and LOD of 7.208 (Tables 1, 2).

A 471-kb haplotype block from Chr6_18844283 to Chr6_19315351 included 7 SNP markers and 17 genes (Figure 3A). Pairwise LD analysis of the imputed SNP data showed that the candidate gene region was from Chr6_18842491 bp to Chr6_19015855 bp (Figure 3B). Seven candidate genes were in the regions, which included polynucleotidyl transferase (Glyma.06G202900 and Glyma.06G203100), polygalacturonase activity (Glyma.06G202600 and Glyma.06G203000), ATP synthase (Glyma.06G203200), and genes without annotation (Glyma.06G202700 and Glyma.06G202800) (Figure 3B).

There were 7 QTL haplotypes in the LD block from Chr6_18844283 to Chr6_19315351 in the natural population that

showed differences in protein content (Supplementary Table 2 and Figure 3C). The haplotypes Hap.B, Hap.C, and Hap.F had higher protein content than other haplotypes. Hap.B had the highest protein content, but no significant difference was observed among Hap.B, Hap.C, Hap.F, and Hap.G (Figure 3C). Further analysis showed that the SNP located at Chr6_19172961 may be more important; varieties carrying Hap.X^{AA} showed higher protein content than Hap.X^{GG} (Figure 3D).

Prediction Accuracy of Seed Protein Content

Prediction accuracy of different SNP densities for seed protein was conducted using 22 significant SNPs resulting from GWAS and 22 to 10,115 random SNPs, respectively. The prediction accuracy ranges from 0.44 to 0.77 using the rrBLUP model and from 0.44 to 0.78 using the BLR model (Figure 4 and Supplementary Table 3). BLR and rrBLUP performed similarly for prediction accuracy; the average prediction accuracy was 0.63 and 0.53, respectively. The prediction accuracy of the 22 SNPs

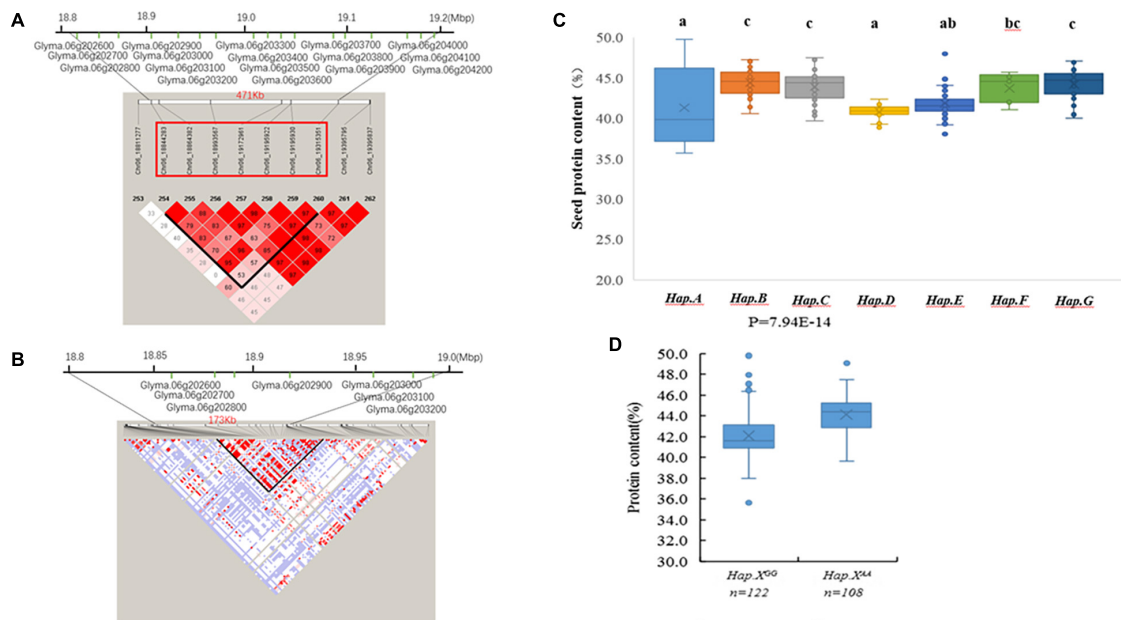


FIGURE 3 | (A) The extent of linkage disequilibrium (LD) in the regions based on pairwise r^2 values. The r^2 values are indicated using the color intensity index. Heatmap showing LD between each pair of markers that passed the Bonferroni threshold in genome-wide association study (GWAS). **(B)** Candidate genes for each single nucleotide polymorphism (SNP) locus. The bottom panel depicts the extent of linkage disequilibrium in the regions based on pairwise r^2 values. The r^2 values are indicated using the color intensity index shown. **(C)** Boxplot of seed protein content based on different genotypes in soybean accessions. **(D)** Boxplot of seed protein content based on Hap.X^{GG} and Hap.X^{AA} phenotypic differences between genotype combinations of the two SNPs.

obtained from GWAS was higher than that of random 22 SNPs and random 250 SNPs (Figure 4 and Supplementary Table 3). Thus, regardless of the GS model, the accuracy of GS was higher when the significant SNPs from GWAS were used. Prediction accuracy for seed protein was increased with higher SNP density. However, there is a minimal difference in prediction accuracy after the SNP number reached 2,000 (Figure 4 and Supplementary Table 3).

The effect of training population size on GS accuracy was also investigated by conducting cross-validation at different folds with 100 replications for each cross-validation (Figure 5 and Supplementary Tables 4, 5). On average, the prediction accuracy of the BLR model was 0.62 using GWAS-derived SNPs and 0.77 using the whole set of SNPs (Figure 5 and Supplementary Table 4). The prediction accuracy of rrBLUP was less than BLR, with 0.5 using GWAS-derived SNPs and 0.77 using the whole set of SNPs (Figure 5 and Supplementary Table 5). Considering average r -value and standardized deviation Sn, sevenfold resulted in a high r -value and low Sn in BLR models and sixfold resulted in a high r -value and low Sn in rrBLUP models.

DISCUSSION

Quantitative Trait Locus Mapping and Candidate Genes Identification for Soybean Seed Protein

Wild soybean with desired traits may improve the yield, quality, and other traits of cultivated soybeans. In this study, we

performed QTL mapping for protein content in a RIL population derived from the cross of cultivated Jidou12 and wild soybean Ye9. Five major stable QTLs were detected on Chr. 6, 8, 15, 17, and 20 using Bayesian IM, SMIM, SMLE, and SMR models in Q-gene and IciMapping. Among these QTLs, we discovered that *qtl-chr6_prot* contributed an average of 25.77 of the phenotypic variance and the positive additive effects of allele were from the cultivated soybean Jidou12. The *qtl-chr6_prot* did not overlap with or was not adjacent to any of the previously reported QTLs for seed protein content. Other QTLs, *qtl-chr8_prot*, *qtl-chr15_prot*, *qtl-chr17_prot*, and *qtl-chr20_prot*, explained an average of 13.99, 9.1, 9.85, and 12.47 of the phenotypic variance, respectively; the positive additive effects of the allele of these QTL were from the wild soybean parent. The QTL *qtl-chr8_prot* (7.27–8.29 Mb) overlapped with the QTLs, as previously reported by Pathan et al. (2013). In addition, the QTL *qtl-chr15_prot* (3.30–4.71 Mb) overlapped with the *qPro15-1* (Zhang et al., 2019) and *qtl-chr17_prot* (12.80–13.81 Mb) with the protein 26-2 (Reinprecht et al., 2006). The position of QTL *qtl-chr20_prot* (26.57–33.51 Mb) was consistent with that of the confirmed QTL *cqPro-20* (Diers et al., 1992; Pandurangan et al., 2012; Vaughn et al., 2014; Sonah et al., 2015; Warrington et al., 2015; Zhang Y. et al., 2018; Fliege et al., 2022). Fliege et al. (2022) concluded that a transposon insertion within the CONSTANS, CO-like, and TOC1 (CCT) domain protein encoded by the *Glyma.20G85100* gene accounted for the high/low seed protein alleles of the *cqSeed* protein-003 QTL (31.74–31.84 Mb).

In the novel QTL region, the *qtl-chr6_prot*, seven candidate genes were identified. Of which, Glyma06G202900 and Glyma06G203100 were annotated as polynucleotidyl transferase,

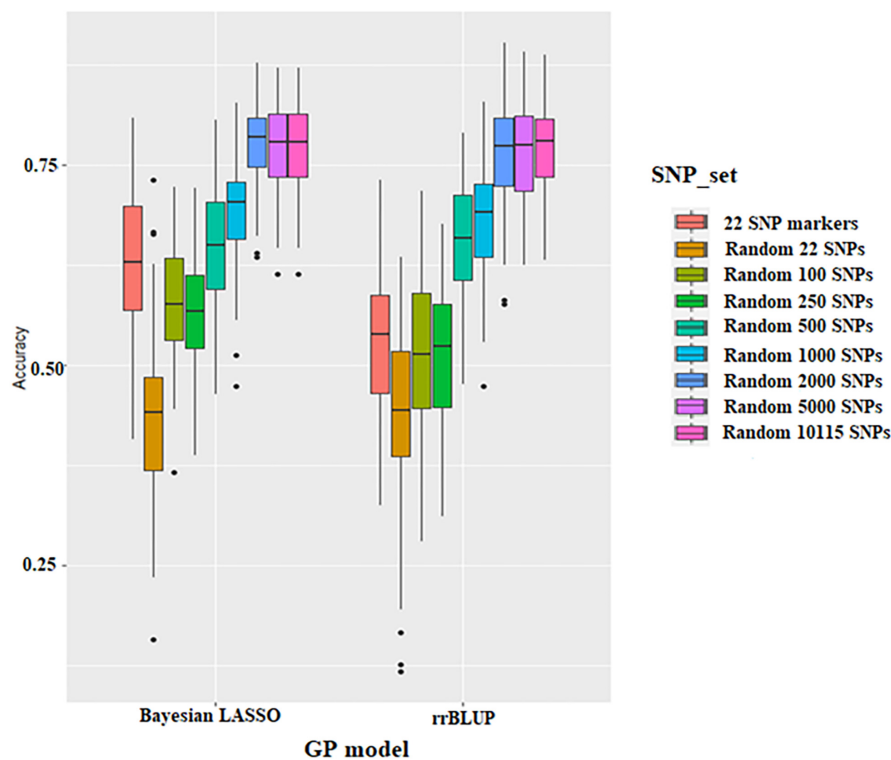


FIGURE 4 | Boxplots show the effect of different SNP density sets on genomic selection in the Bayesian Lasso Regression (BLR) model and ridge regression best linear unbiased prediction (rrBLUP) models.

ribonuclease H-like superfamily protein, which were homologous to the AT5G61090 gene in *Arabidopsis*. The protein encoded by the AT5G61090 had an RNA–DNA hybrid ribonuclease activity (Stoppel and Meurer, 2012). Glyma06G202600 was annotated as plasmodesmata callose-binding protein 3, homologous to AT1G18650 with callose-binding activity and the regulating intercellular trafficking in *Arabidopsis* (Simpson et al., 2009). Glyma06G203000 was annotated as a pectin lyase-like superfamily protein homologous to AT3G07820 with a polygalacturonase activity in *Arabidopsis* (Kim et al., 2006). Glyma06G203200 was annotated as a gamma subunit of Mt ATP synthase, homologous to AT2G33040, one of mitochondrial (mt) ATP synthesis subunits. Reduced expression of these subunits of the mt ATP synthase was proposed to disturb cellular redox states (Robison et al., 2009).

Genomic Selection in Soybean

Genomic selection overcomes the problems of traditional breeding methods and MAS selection and provides a new way for the selection of quantitative traits controlled by genes with minor effects. GS allows for the estimation of the effects of all the markers across the genome. These effects can be used to predict the performance of lines (Meuwissen et al., 2001). Since the target trait phenotype of an individual is predicted using the GS model, the materials could be screened and selected before planting, thus reducing costs and improving breeding efficiency (Heslot et al., 2012; Longin et al., 2015; Spindel et al., 2015). Matei

et al. (2018) showed that the selection cycle for yield and seed weight can be significantly shortened using GS.

So far, the GS study has been mainly conducted on maize, wheat, and rice. The GS study in soybean remains limited. In 2013, Shu performed GS for 100-seed weight and reported a prediction accuracy of 0.904 (Shu et al., 2013). Subsequent GS showed accuracy for soybean cyst nematode (SCN) was 0.59–0.67 (Bao et al., 2014) and 0.64 for soybean yield (Jarquín et al., 2014).

The GS was performed on amino acid concentration (Qin et al., 2019), soybean chlorophyll content, soybean cyst nematode tolerance (Ravelombola et al., 2019), yield, and yield-related traits, such as maturity, plant height, and 100-seed weight (Ravelombola et al., 2021). These studies have shown the feasibility of GS for soybean yield and quality-related traits (Matei et al., 2018; Stewart-Brown et al., 2019).

However, few reports have focused on the GS of seed protein in soybean. Stewart-Brown et al. (2019) evaluated the potential of GS for soybean seed protein using 483 elite breeding lines from 26 biparentals and reported the predictive abilities of 0.81 in all the populations, 0.55 across populations, and 0.60 within each biparental population. Duhnen et al. (2017) compared genomic prediction accuracy of seed protein obtained using models calibrated across or within two subpopulations: early lines and late lines. The results showed that calibrations within subpopulations were more efficient. Five Bayesian models were also compared with Genomic best linear unbiased prediction (GBLUP) and did not show improved prediction accuracy. In this

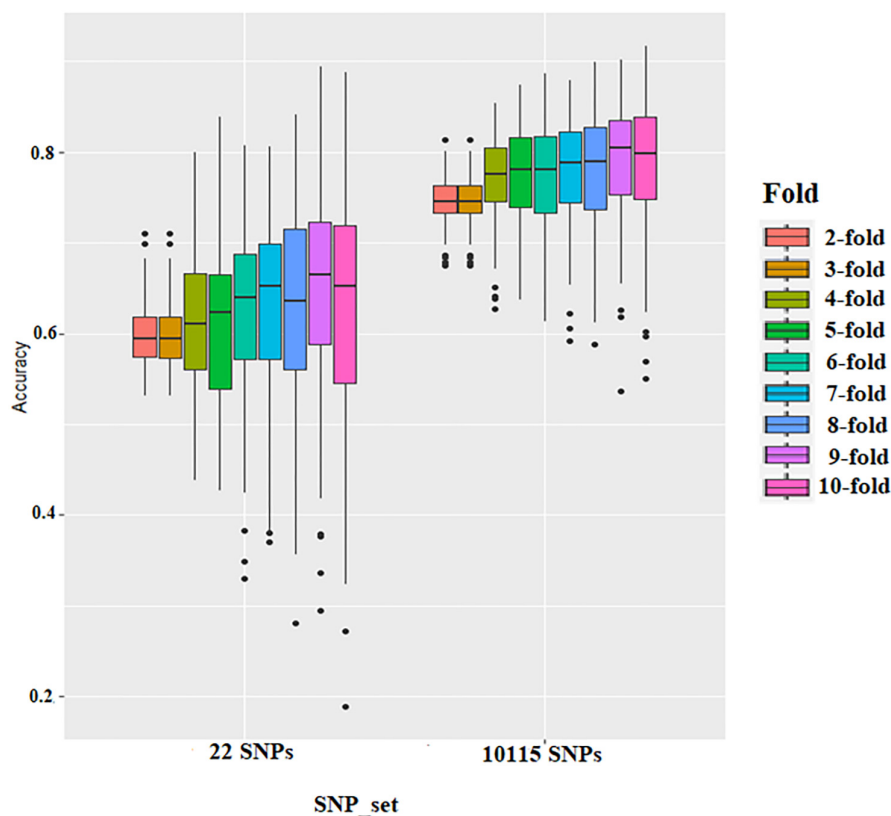


FIGURE 5 | Boxplots show the effect of training population size on genomic selection accuracy by conducting cross-validation at different folds with 100 replications for each cross-validation fold using rrBLUP.

study, we performed GS based on different SNP sets, different training population sizes, and statistical models. The results showed that the use of GWAS-derived SNPs for conducting GS significantly improved the accuracy of prediction, which was consistent with the results reported by Qin et al. (2019). The model selection criteria, SNP sets, and population training size were critical factors when conducting a GS, as reported in previous studies (Ravelombola et al., 2019, 2020, 2021). Those studies had demonstrated that 1,000–2,000 genome-wide markers across all the lines/accessions were needed to reach maximum efficiency of genomic prediction in the populations, increasing marker density that would not improve prediction efficiency (Poland et al., 2012; Bao et al., 2014; Zhang J. et al., 2016; Song et al., 2020). This study showed that there was a minimal difference in prediction accuracy after the SNP number reached 2,000 for seed protein content.

CONCLUSION

This study reported mapping and GS for seed protein content. Molecular markers associated with seed protein content were identified in RIL and natural populations and a novel QTL for seed protein content was detected and mapped on Chr. 6 in both populations. In addition, seven candidate genes that were

related to seed protein content were identified. This is one of a few reports investigating seed protein content using RILs derived from cultivated and wild soybean crosses. Our results showed that GS accuracy was dependent on the SNP set and training population size; a set of GWAS-derived SNPs could increase GS accuracy. No significant GS accuracy difference was observed between rrBLUP and BL models. The results demonstrated the potential of using GS to improve soybean seed protein content.

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

JQ and AS: data curation. JQ, CY, and LY: funding acquisition. FW, QZ, and TZ: investigation. JQ, AS, and WR: methodology. MZ, LY, and CY: project administration. AS: software. FW and TZ: validation. JQ: writing – original draft preparation. JQ, QS, and AS: writing - review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by: (1) the National Natural Science Foundation of China (32072092); (2) the Basic Research Funds of Hebei Academy of Agriculture and Forestry Sciences (2021060205); (3) the Special Innovation Program of Hebei Academy of Agriculture and Forestry Sciences (2022KJCXZX-LYS-6); (4) the S&T Program of Hebei, Soybean Modern Seed Industry Science and Technology Innovation Team (21326313D); (5) the Hebei Natural Science Foundation (2020301020); and (6) the China

Agriculture Research System of The Ministry of Finance (MOF) and The Ministry of Agriculture and Rural Affairs (MARA) (CARS-04).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.882732/full#supplementary-material>

REFERENCES

- Bandillo, N., Jarquin, D., Song, Q., Nelson, R. L., Cregan, P., Specht, J., et al. (2015). A population structure and genome-wide association analysis on the USDA soybean germplasm collection. *Plant Genome* 8, 1–13. doi: 10.3835/plantgenome2015.04.0024
- Bao, Y., Vuong, T., Meinhardt, C., Tiffin, P., Denny, R., Chen, S., et al. (2014). Potential of association mapping and genomic selection to explore PI 88788 derived soybean cyst nematode resistance. *Plant Genome* 7, 2840–2854.
- 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
- Brummer, E., Graef, G., Orf, J., Wilcox, J., and Shoemaker, R. (1997). Mapping QTL for seed protein and oil content in eight soybean populations. *Crop Sci.* 37, 370–378.
- Chapman, A., Pantalone, V., Ustun, A., Allen, F., Landau-Ellis, D., Triganio, R., et al. (2003). Quantitative trait loci for agronomic and seed quality traits in an F₂ and F₄: 6 soybean population. *Euphytica* 129, 387–393.
- Collard, B. C., Jahufer, M. Z. Z., Brouwer, J. B., and Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. *Euphytica* 142, 169–196.
- Csanádi, G., Vollmann, J., Stift, G., and Lelley, T. (2001). Seed quality QTLs identified in a molecular map of early maturing soybean. *Theor. Appl. Genet.* 103, 912–919.
- Diers, B. W., Keim, P., Fehr, W., and Shoemaker, R. (1992). RFLP analysis of soybean seed protein and oil content. *Theor. Appl. Genet.* 83, 608–612. doi: 10.1007/BF00226905
- Duhnen, A., Gras, A., Teyssière, S., Romestant, M., Claustres, B., Daydé, J., et al. (2017). Genomic selection for yield and seed protein content in soybean: a study of breeding program data and assessment of prediction accuracy. *Crop Sci.* 57, 1325–1337.
- Earl, D. A., and vonHoldt, B. M. (2012). Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv. Genet. Resour.* 4, 359–361.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* 6:e19379. doi: 10.1371/journal.pone.0019379
- Endelman, J. B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome* 4, 250–255.
- Fliege, C. E., Ward, R. A., Vogel, P., Nguyen, H., Quach, T., Guo, M., et al. (2022). Fine mapping and cloning of the major seed protein QTL on soybean chromosome 20. *Plant J.* 110, 114–128. doi: 10.1111/tpj.15658
- Gangurde, S. S., Wang, H., Yaduru, S., Pandey, M. K., Fountain, J. C., Chu, Y., et al. (2020). Nested-association mapping (NAM)-based genetic dissection uncovers candidate genes for seed and pod weights in peanut (*Arachis hypogaea*). *Plant Biotechnol. J.* 18, 1457–1471. doi: 10.1111/pbi.13311
- Hacisalihoglu, G., Burton, A. L., Gustin, J. L., Eker, S., Asikli, S., Heybet, E. H., et al. (2018). Quantitative trait loci associated with soybean seed weight and composition under different phosphorus levels. *J. Integr. Plant Biol.* 60, 232–241. doi: 10.1111/jipb.12612
- Heslot, N., Yang, H. P., Sorrells, M. E., and Jannink, J. L. (2012). Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52, 146–160.
- Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1. doi: 10.1186/1471-2164-15-1
- Hyten, D. L., Pantalone, V. R., Sams, C., Saxton, A., Landau-Ellis, D., Stefaniak, T., et al. (2004). Seed quality QTL in a prominent soybean population. *Theor. Appl. Genet.* 109, 552–561. doi: 10.1007/s00122-004-1661-5
- Jarquin, D., Kocak, K., Posadas, L., Hyma, K., Jedlicka, J., Graef, G., et al. (2014). Genotyping by sequencing for genomic prediction in a soybean breeding population. *BMC Genomics* 15:740. doi: 10.1186/1471-2164-15-740
- Joehanes, R., and Nelson, J. C. (2008). QGene 4.0, an extensible Java QTL-analysis platform. *Bioinformatics* 24, 2788–2789. doi: 10.1093/bioinformatics/btn523
- Kaeuffer, R., Réale, D., Coltman, D., and Pontier, D. (2007). Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity* 99, 374–380. doi: 10.1038/sj.hdy.6801010
- Kim, J., Shiu, S. H., Thoma, S., Li, W. H., and Patterson, S. E. (2006). Patterns of expansion and expression divergence in the plant polygalacturonase gene family. *Genome Biol.* 7, 1–14. doi: 10.1186/gb-2006-7-9-r87
- Lam, H.-M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Leamy, L. J., Zhang, H., Li, C., Chen, C. Y., and Song, B. H. (2017). A genome-wide association study of seed composition traits in wild soybean (*Glycine soja*). *BMC Genomics* 18:18. doi: 10.1186/s12864-016-3397-4
- Lee, S., Van, K., Sung, M., Nelson, R., Lamantia, J., Mchale, L. K., et al. (2019). Genome-wide association study of seed protein, oil and amino acid contents in soybean from maturity groups I to IV. *Theor. Appl. Genet.* 132, 1639–1659. doi: 10.1007/s00122-019-03304-5
- Legarra, A., Robert-Granie, C., Croiseau, P., Guillaume, F., and Fritz, S. (2011). Improved Lasso for genomic selection. *Genet. Res.* 93, 77–87. doi: 10.1017/S0016672310000534
- Li, D., Zhao, X., Han, Y., Li, W., and Xie, F. (2019). Genome-wide association mapping for seed protein and oil contents using a large panel of soybean accessions. *Genomics* 111, 90–95. doi: 10.1016/j.ygeno.2018.01.004
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27, 2987–2993. doi: 10.1093/bioinformatics/btr509
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352
- Li, S., Xu, H., Yang, J., and Zhao, T. (2019). Dissecting the genetic architecture of seed protein and oil content in soybean from the Yangtze and Huaihe River valleys using multi-locus genome-wide association studies. *Int. J. Mol. Sci.* 20:3041. doi: 10.3390/ijms20123041
- Liang, H. Z., Yu, Y. L., Wang, S. F., Yun, L., Wang, T. F., Wei, Y. L., et al. (2010). QTL mapping of isoflavone, oil and protein contents in soybean (*Glycine max* L. Merr.). *Agr. Sci. China* 9, 1108–1116.
- Liu, X., Wang, H., Wang, H., Guo, Z., Xu, X., Liu, J., et al. (2018). Factors affecting genomic selection revealed by empirical evidence in maize. *Crop J.* 6, 341–352.
- Longin, C. F. H., Mi, X., and Würschum, T. (2015). Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for

- line and hybrid breeding. *Theor. Appl. Genet.* 128, 1297–1306. doi: 10.1007/s00122-015-2505-1
- Matei, G., Woyann, L. G., Milioli, A. S., De Bem Oliveira, I., Zdzarski, A. D., Zanella, R., et al. (2018). Genomic selection in soybean: accuracy and time gain in relation to phenotypic selection. *Mol. Breed.* 38, 1–13.
- Meng, L., Li, H. H., Zhang, L. Y., and Wang, J. K. (2015). QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J.* 3, 269–283. doi: 10.1016/j.cj.2015.01.001
- Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829. doi: 10.1093/genetics/157.4.1819
- Nichols, D., Glover, K., Carlson, S., Specht, J., and Diers, B. (2006). Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. *Crop Sci.* 46, 834–839.
- Pandurangan, S., Pajak, A., Molnar, S. J., Cober, E. R., Dhaubhadel, S., Hernández-Sebastiá, C., et al. (2012). Relationship between asparagine metabolism and protein concentration in soybean seed. *J. Exp. Bot.* 63, 3173–3184. doi: 10.1093/jxb/ers039
- Pathan, S. M., Vuong, T., Clark, K., Lee, J. D., Shannon, J. G., Roberts, C. A., et al. (2013). Genetic mapping and confirmation of quantitative trait loci for seed protein and oil contents and seed weight in soybean. *Crop Sci.* 53, 765–774.
- Patil, G., Mian, R., Vuong, T., Pantalone, V., Song, Q., Chen, P., et al. (2017). Molecular mapping and genomics of soybean seed protein: a review and perspective for the future. *Theor. Appl. Genet.* 130, 1975–1991. doi: 10.1007/s00122-017-2955-8
- Pérez, P., and de los Campos, G. (2014). Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198, 483–495. doi: 10.1534/genetics.114.164442
- Poland, J. A., Endelman, J., Dawson, J., Rutkoski, J., Wu, S., Manes, Y., et al. (2012). Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome* 5, 103–113.
- Qi, Z., Hou, M., Han, X., Liu, C., Jiang, H., Xin, D., et al. (2014). Identification of quantitative trait loci (QTL s) for seed protein concentration in soybean and analysis for additive effects and epistatic effects of QTL s under multiple environments. *Plant Breed.* 133, 499–507.
- Qin, J., Shi, A., Song, Q., Li, S., Wang, F., Cao, Y., et al. (2019). Genome wide association study and genomic selection of amino acid concentrations in soybean seeds. *Front. Plant Sci.* 10:1445. doi: 10.3389/fpls.2019.01445
- Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2019). Genome-wide association study and genomic selection for soybean chlorophyll content associated with soybean cyst nematode tolerance. *BMC Genomics* 20:904. doi: 10.1186/s12864-019-6275-z
- Ravelombola, W. S., Qin, J., Shi, A., Nice, L., Bao, Y., Lorenz, A., et al. (2020). Genome-wide association study and genomic selection for tolerance of soybean biomass to soybean cyst nematode infestation. *PLoS One* 15:e0235089. doi: 10.1371/journal.pone.0235089
- Ravelombola, W., Qin, J., Shi, A., Song, Q., Yuan, J., Wang, F., et al. (2021). Genome-wide association study and genomic selection for yield and related traits in soybean. *PLoS One* 16:e0255761. doi: 10.1371/journal.pone.0255761
- Reinprecht, Y., Poysa, V. W., Yu, K., Rajcan, I., Ablett, G. R., and Pauls, K. P. (2006). Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (*Glycine max* (L.) Merrill) germplasm. *Genome* 49, 1510–1527. doi: 10.1139/g06-112
- Robison, M. M., Ling, X., Smid, M. P., Zarei, A., and Wolyn, D. J. (2009). Antisense expression of mitochondrial ATP synthase subunits OSCP (ATP5) and γ (ATP3) alters leaf morphology, metabolism and gene expression in Arabidopsis. *Plant Cell Physiol.* 50, 1840–1850. doi: 10.1093/pcp/pcp125
- Sall, J., Stephens, M. L., Lehman, A., and Loring, S. (2017). *JMP Start Statistics: A Guide to Statistics and Data Analysis Using JMP*. Cary, NC: SAS Institute.
- 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
- Sebolt, A., Shoemaker, R., and Diers, B. (2000). Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. *Crop Sci.* 40, 1438–1444.
- Shu, Y., Yu, D., Wang, D., Bai, X., Zhu, Y., and Guo, C. (2013). Genomic selection of seed weight based on low-density SCAR markers in soybean. *Genet. Mol. Res.* 12, 2178–2188. doi: 10.4238/2013.July.3.2
- Simpson, C., Thomas, C., Findlay, K., Bayer, E., and Maule, A. J. (2009). An Arabidopsis GPI-anchor plasmodesmal neck protein with callose binding activity and potential to regulate cell-to-cell trafficking. *Plant Cell* 21, 581–594. doi: 10.1105/tpc.108.060145
- Sonah, H., O'donoghue, L., Cober, E., Rajcan, I., and Belzile, F. (2015). Identification of loci governing eight agronomic traits using a GBS-GWAS approach and validation by QTL mapping in soya bean. *Plant Biotechnol. J.* 13, 211–221. doi: 10.1111/pbi.12249
- Song, Q., Hyten, D. L., Jia, G., Quigley, C. V., Fickus, E. W., Nelson, R. L., et al. (2013). Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. *PLoS One* 8:e54985. doi: 10.1371/journal.pone.0054985
- Song, Q., Marek, L., Shoemaker, R., Lark, K., Concibido, V., Delannay, X., et al. (2004). A new integrated genetic linkage map of the soybean. *Theor. Appl. Genet.* 109, 122–128. doi: 10.1007/s00122-004-1602-3
- Song, Q., Yan, L., Quigley, C., Fickus, E., Wei, H., Chen, L., et al. (2020). Soybean BARCSoySNP6K: an assay for soybean genetics and breeding research. *Plant J.* 104, 800–811. doi: 10.1111/tpj.14960
- Spindel, J., Begum, H., Akdemir, D., Virk, P., Collard, B., Redona, E., et al. (2015). Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. *PLoS Genet.* 11:e1004982. doi: 10.1371/journal.pgen.1004982
- Stewart-Brown, B. B., Song, Q., Vaughn, J. N., and Li, Z. (2019). Genomic selection for yield and seed composition traits within an applied soybean breeding program. *G3* 9, 2253–2265. doi: 10.1534/g3.118.200917
- Stoppel, R., and Meurer, J. (2012). The cutting crew—ribonucleases are key players in the control of plastid gene expression. *J. Exp. Bot.* 63, 1663–1673. doi: 10.1093/jxb/err401
- Teng, W., Li, W., Zhang, Q., Wu, D., Zhao, X., Li, H., et al. (2017). Identification of quantitative trait loci underlying seed protein content of soybean including main, epistatic, and QTL \times environment effects in different regions of Northeast China. *Genome* 60, 649–655. doi: 10.1139/gen-2016-0189
- Van Ooijen, J. W. (2006). *JoinMap 4: Software for the Calculation of Genetic Linkage Maps in Experimental Populations*. Wageningen: Kyazma B.V.
- Vaughn, J. N., Nelson, R. L., Song, Q., Cregan, P. B., and Li, Z. (2014). The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. *G3* 4, 2283–2294. doi: 10.1534/g3.114.013433
- Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Yu, Y. C., et al. (2020). Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. *Natl. Sci. Rev.* 7, 1776–1786. doi: 10.1093/nsr/nwaa110
- Wang, X., Jiang, G. L., Green, M., Scott, R. A., Song, Q., Hyten, D. L., et al. (2014). Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. *Mol. Genet. Genomics* 289, 935–949. doi: 10.1007/s00438-014-0865-x
- Warrington, C. V., Abdel-Haleem, H., Hyten, D., Cregan, P., Orf, J., Killam, A., et al. (2015). QTL for seed protein and amino acids in the Benning \times Danbaekkong soybean population. *Theor. Appl. Genet.* 128, 839–850.
- Whiting, R. M., Torabi, S., Lukens, L., and Eskandari, M. (2020). Genomic regions associated with important seed quality traits in food-grade soybeans. *BMC Plant Biol.* 20:485. doi: 10.1186/s12870-020-02681-0
- Wolf, W. J. (1970). Soybean proteins. Their functional, chemical, and physical properties. *J. Agric. Food Chem.* 18, 969–976.
- Xu, Y., and Crouch, J. H. (2008). Marker-assisted selection in plant breeding: from publications to practice. *Crop Sci.* 48, 391–407.
- Xu, Y., Lu, Y., Xie, C., Gao, S., Wan, J., and Prasanna, B. M. (2012). Whole-genome strategies for marker-assisted plant breeding. *Mol. Breeding* 29, 833–854.
- Yan, L., Jiang, C. Z., Yu, X. H., Yang, C. Y., and Zhang, M. C. (2008). Development and reliability of near infrared spectroscopy (NIS) models of protein and oil content in soybean. *Soybean Sci.* 27, 833–837.
- Yang, J., Hu, C., Hu, H., Yu, R., Xia, Z., Ye, X., et al. (2008). QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. *Bioinformatics* 24, 721–723. doi: 10.1093/bioinformatics/btm494
- Yao, Y., You, Q., Duan, G., Ren, J., Chu, S., Zhao, J., et al. (2020). Quantitative trait loci analysis of seed oil content and composition of wild and cultivated soybean. *BMC Plant Biol.* 20:51. doi: 10.1186/s12870-019-2199-7

- Yu, J., Pressoir, G., Briggs, W. H., Vroh, Bi I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhang, A., Wang, H., Beyene, Y., Semagn, K., Liu, Y., Cao, S., et al. (2017). Effect of trait heritability, training population size and marker density on genomic prediction accuracy estimation in 22 bi-parental tropical maize populations. *Front. Plant Sci.* 8:1916. doi: 10.3389/fpls.2017.01916
- Zhang, D., Lü, H., Chu, S., Zhang, H., Zhang, H., Yang, Y., et al. (2017). The genetic architecture of water-soluble protein content and its genetic relationship to total protein content in soybean. *Sci. Rep.* 7:5636. doi: 10.1038/s41598-017-04685-7
- Zhang, J., Song, Q., Cregan, P. B., and Jiang, G. L. (2016). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129, 117–130. doi: 10.1007/s00122-015-2614-x
- Zhang, J., Wang, X., Lu, Y., Bhusal, S. J., Song, Q., Cregan, P. B., et al. (2018). Genome-wide scan for seed composition provides insights into soybean quality improvement and the impacts of domestication and breeding. *Mol. Plant* 11, 460–472. doi: 10.1016/j.molp.2017.12.016
- Zhang, T., Wu, T., Wang, L., Jiang, B., Zhen, C., Yuan, S., et al. (2019). A combined linkage and GWAS analysis identifies QTLs linked to soybean seed protein and oil content. *Int. J. Mol. Sci.* 20, 1–19. doi: 10.3390/ijms20235915
- Zhang, Y., He, J., Meng, S., Liu, M., Xing, G., Li, Y., et al. (2018). Identifying QTL–allele system of seed protein content in Chinese soybean landraces for population differentiation studies and optimal cross predictions. *Euphytica* 214, 1–17.

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 Qin, Wang, Zhao, Shi, Zhao, Song, Ravelombola, An, Yan, Yang 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.



Analysis of Physiological Variations and Genetic Architecture for Photosynthetic Capacity of Japanese Soybean Germplasm

Mohammad Jan Shamim¹, Akito Kaga^{2*}, Yu Tanaka^{1*}, Hiroshi Yamatani² and Tatsuhiko Shiraiwa¹

¹Laboratory of Crop Science, Division of Agronomy and Horticultural Science, Graduate School of Agriculture, Kyoto University, Kyoto, Japan, ²Institute of Crop Science, National Agriculture and Food Research Organization, Tsukuba, Japan

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Wenqiang Yang,
Key Laboratory of Photobiology,
Institute of Botany (CAS), China
Shuoqi Chang,
Hunan Academy of Agricultural
Sciences, China

*Correspondence:

Yu Tanaka
tanaka.yu.2s@kyoto-u.ac.jp
Kaga Akito
kaga@affrc.go.jp

Specialty section:

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

Received: 01 April 2022

Accepted: 01 June 2022

Published: 29 June 2022

Citation:

Shamim MJ, Kaga A, Tanaka Y,
Yamatani H and Shiraiwa T (2022)
Analysis of Physiological Variations
and Genetic Architecture for
Photosynthetic Capacity of Japanese
Soybean Germplasm.
Front. Plant Sci. 13:910527.
doi: 10.3389/fpls.2022.910527

The culmination of conventional yield improving parameters has widened the margin between food demand and crop yield, leaving the potential yield productivity to be bridged by the manipulation of photosynthetic processes in plants. Efficient strategies to assess photosynthetic capacity in crops need to be developed to identify suitable targets that have the potential to improve photosynthetic efficiencies. Here, we assessed the photosynthetic capacity of the Japanese soybean mini core collection (GmJMC) using a newly developed high-throughput photosynthesis measurement system “MIC-100” to analyze physiological mechanisms and genetic architecture underpinning photosynthesis. K-means clustering of light-saturated photosynthesis (A_{sat}) classified GmJMC accessions into four distinct clusters with Cluster2 comprised of highly photosynthesizing accessions. Genome-wide association analysis based on the variation of A_{sat} revealed a significant association with a single nucleotide polymorphism (SNP) on chromosome 17. Among the candidate genes related to photosynthesis in the genomic region, variation in expression of a gene encoding G protein alpha subunit 1 (GPA1) showed a strong correlation ($r=0.72$, $p<0.01$) with that of A_{sat} . Among GmJMC accessions, GmJMC47 was characterized by the highest A_{sat} , stomatal conductance (g_s), stomatal density (S_{Density}), electron transfer rate (ETR), and light use efficiency of photosystem II (F_v'/F_m') and the lowest non-photochemical quenching [$NPQ(t)$], indicating that GmJMC47 has greater CO_2 supply and efficient light-harvesting systems. These results provide strong evidence that exploration of plant germplasm is a useful strategy to unlock the potential of resource use efficiencies for photosynthesis.

Keywords: soybean core collection, gas exchange, K-means clustering analysis, GWAS, RNA expression

INTRODUCTION

It is estimated that the current rate of global crop production needs to be doubled to meet the demands of the rising population, changing diet, and requirement for biofuel (Tilman et al., 2011). Although crop seed yield has increased in many parts of the world, its stagnation in an estimated 24–27% of the areas, where maize, rice, wheat, and soybean are grown,

underscores the challenge to address the rising demands for agriculture (Ray et al., 2012). Climate change including, rising temperature, drought, flooding, and air pollution further aggravate this challenge (Christensen and Hewitson, 2007). With the decrease in arable land and resources, developing novel strategies to maintain sustainable plant production for the food and energy demand of humanity is essential (Ray et al., 2012; Maurino and Weber, 2013). Conventional breeding and selection of cultivars have been used to improve yield and it is estimated that rice and wheat have acquired ~1% annual gain in yield potential in the last 30 years (Stapper and Fischer, 1990; Khush, 2001). However, recent studies have reported the exhaustion of the traits that led to a remarkable increase in crop seed yield during and after the green revolution (Zhu et al., 2010). An effective way of producing sufficient food, practicing sustainability, and avoiding the exploitation of natural resources is suggested to be the manipulation of photosynthesis (Janssen et al., 2014).

Improving photosynthetic capabilities in plants, however, is a complex process and requires the removal of inefficiencies from the photosynthetic machinery. Photosynthesis is majorly limited by biochemical and diffusional factors and several targets have been proposed for genetic manipulation to increase photosynthetic capacity. For instance, copious quantities of Rubisco that catalyze two competitive processes of CO₂ fixation and oxygenation are required for better photosynthetic capacity. Changes in the quantity, specificity, or affinity of CO₂ in Rubisco can boost photosynthesis (Parry et al., 2012). Reduction in transient photoinhibition, increase in ribulose 1,5-bisphosphate regeneration capacity, and increase in specificity of Rubisco for CO₂ have been considered ideal targets for the genetic manipulation of light-saturated photosynthesis in C3 crops (Lawlor et al., 1999). The overexpression of chloroplast NAD kinase (NADK2) was observed to play an important role in photosynthetic electron transfer thus improving photosynthesis (Takahara et al., 2010). Similarly, a study on transgenic *Arabidopsis thaliana* revealed that increased stomatal density (maximum 372%) increased the photosynthetic rate by 30 percent (Tanaka et al., 2013). The genetic manipulation of *g_s* in *slac1* mutant (lacking stomatal anion channel protein that controls stomatal closure) revealed that a higher *g_s* enhances the rate of photosynthesis as well as the ratio of internal to external CO₂ concentrations (Kusumi et al., 2012). Mounting evidence suggests that the genetic manipulation of rate-limiting processes enhances photosynthesis. The problem, however, is that a large part of plant germplasms remains unexplored. Harnessing natural variations in photosynthesis, therefore, is an ideal route toward removing bottlenecks in photosynthesis.

The exploration of plant germplasms for identifying promising accessions and elucidating the association between genetic architecture and photosynthesis is important to enhance photosynthesis. Kromdijk and Long (2016) stated that there is a large gap between the number of germplasms preserved in gene banks and those used in crop breeding. They argue that delayed photosynthetic improvement is largely due to the plant germplasm that remains unexplored. The exploration of core collections is, therefore, essential for identifying new targets

to use in crop improvement (Upadhyaya, 2015). On the other hand, studies have been conducted to assess the possibility of enhancing photosynthesis through genomics and gene modification. Lopez et al. (2019) reported a high genetic correlation ($r=0.80$) between seed yield and photosynthesis in the US soybean nested association map (NAM) panel. They also found single nucleotide polymorphisms (SNPs) which were associated with photosynthesis. Vieira et al. (2006) identified some quantitative trait loci (QTLs) on chromosomes 10 and 16 that were associated with light-saturated rates of photosynthesis. Simkin et al. (2015) indicated that increasing the levels of sedoheptulose-1,7-bisphosphatase, fructose 1,6-bisphosphate aldolase, and the cyanobacterial putative-inorganic carbon transporter b (ictB) improved photosynthetic CO₂ fixation (12–19% in mature plants), leaf area, biomass, and yield in transgenic tobacco. Xiao et al. (2020) found that heterologous expression of the highly linked genes (*PtoPsbX1*) in their gene expression network significantly improved photosynthesis in *A. thaliana*. These findings collectively suggest that exploring the plant germplasms for natural variations in photosynthetic capacity and studying the genetic architecture underlying photosynthesis is an ideal strategy for improving photosynthetic efficiency in plants.

Soybean is a major commodity crop (Hay et al., 2017) and an indispensable part of the Japanese diet; it is also said to increase longevity among Japanese people (Gabriel et al., 2018). Japanese soybean cultivars, however, are less productive than the US cultivars, as the latter have higher radiation use efficiency (RUE), greater photosynthetic capacity, and dry matter production (Kawasaki et al., 2016). GmJMC (JMC hereinafter) is a small but diverse germplasm set named mini core collection that represents a major part of geographical and agro-morphological trait variations of Japanese soybean landraces (Kaga et al., 2012). It is consisted of 96 accessions selected from 1,603 soybean accessions based on genetic variation. However, JMC has never been evaluated for its genetic variations in leaf photosynthetic capacity. Considering the geographical and agro-morphological characteristics of this collection, we assume that studying leaf photosynthetic capacity would divulge important information and provide suitable targets for resource use efficiencies. This study aimed at assessing the photosynthetic capacity of JMC accessions using a newly developed high-throughput photosynthesis measurement system, MIC-100, by analyzing physiological variations and genetic architecture underlying photosynthetic discrepancies and identifying suitable targets to improve photosynthetic efficiency.

MATERIALS AND METHODS

Plant Materials

All germplasms were obtained from National Agricultural and Food Research Organization (NARO) Genebank (Japan). Two experiments were conducted for this study. The first experiment was aimed at assessing phenotypic variations in *A_{sat}*. We selected 90 accessions from the JMC and sowed

their seeds on 18 June 2019; among which 16 accessions revealed lodging and poor seed quality. The remaining 74 accessions were cultivated again on 20 June 2020 using a randomized complete design (RCD). Ten plants were sown in each plot.

The second experiment was conducted in 2021 to measure various parameters related to photosynthetic capacity in 10 representative accessions (indicated by white dots inside markers) and investigate the mechanisms underlying it. We selected seven-candidate accessions (JMC13, JMC16, JMC37, JMC43, JMC47, JMC49) from the best photosynthesizing Cluster2; two accessions (JMC56, and JMC112) from Cluster1; one accession (JMC25) from Cluster3; and a Chinese accession Peking (named “PE” as a check for having higher A_{sat} and g_s). JMC25 and JMC112 are two photosynthetically elite cultivars, named Enrei (“En”) and Fukuyutaka (“Fu”), respectively. These accessions were cultivated in a randomized complete block design (RCBD) with two replicates.

All experiments were conducted in a field at the Laboratory of Crop Science, Graduate School of Agriculture, Kyoto University (35°0.2' N, 135° 47' E). The sowing density was 9.52 plants m^{-2} . The space between rows and between plants in a row was 0.7 m and 0.15 m, respectively. We applied 3:10:10 gm^{-2} of $\text{N:P}_2\text{O}_5:\text{K}_2\text{O}$ fertilizers during sowing. Three seeds were planted per hill of which only one was retained after unifoliate leaves had fully expanded. Irrigation and pesticide application were done regularly.

Photosynthetic Phenotyping

We used a high-throughput portable close-chamber infrared gas analyzer, MIC-100 (MASA International Cooperation, Japan), to analyze light-saturated photosynthetic rates (A_{sat}) in 2019 and 2020. This system requires less than 30 s per sample whereas a standard LI-6800 achieves stability in at least 45 s for the “Fast Survey Measurements.” MIC100 explains around 95% variations in A_{sat} measured by a standard LI-6800 system (LI-COR, United States) therefore, making it a high-throughput efficient system (LI-COR Biosciences, 2021; Tanaka et al., 2021). Moreover, our observations in the field revealed that MIC-100 can measure A_{sat} up to sevenfold faster than LI-6800. Measurements were conducted for two consecutive years: three measurements in 2019 (5 August, 13 August, and 18 August) and five measurements in 2020 (06 August, 14 August, 21 August, 27 August, and 31 August). The data were collected from 9:00 AM to 1:30 PM from a central leaflet of a fully expanded recently matured trifoliolate from flower initiation stage (R1) to seed initiation stage (R5). We were able to measure A_{sat} in 360 and 240 leaves on each measurement day for 2019 and 2020, respectively which collectively sums up to 2,280 samples. Only one plant was sampled at a time so that A_{sat} is distributed throughout the survey time to compensate for the diurnal privilege of accessions. Light intensity inside the cuvette was $1,200 \mu\text{mol m}^{-2} \text{s}^{-1}$ which is the maximum capacity of this system. Temperature is automatically recorded at the end of each experiment and did not exceed 36°C. Ambient CO_2 and humidity conditions were used for the experiment,

with the initial CO_2 concentration set at (370–390 ppm) and the final concentration set 20 points lower than the initial concentration.

In 2021, an LI-6800 (LI-COR, United States) was used on 22 July, 30 July, and 11 August for gas exchange and chlorophyll fluorescence measurements. These parameters were measured in four randomly selected plants from each plot from 9:00 AM to 11:00 AM. The temperature inside the chamber cuvette was 33°C. Reference level CO_2 was set at $400 \mu\text{mol mol}^{-1}$ using CO_2 cartridges, with an airflow rate of $500 \mu\text{mol s}^{-1}$. Light intensity was set to $1,200 \mu\text{mol m}^{-2} \text{s}^{-1}$ with 10% blue light and relative humidity was set at 60%. We calculated NPQ(t), as an alias to NPQ, following Tietz et al. (2017).

K-Means Clustering Analysis

Before performing K-means clustering analysis, we used T-distributed Stochastic Neighbor Embedding (t-SNE) to reduce the dimensionality of the data (A_{sat} was measured 8 times throughout 2019 and 2020). The data were standardized so that each measurement had a mean of zero and a standard deviation of one. The number of components in t-SNE was set to two and we gave it 1,000 iterations. The output data were used in the K-means clustering algorithm with four centroids.

Genome-Wide Association Study

The SNP dataset obtained for 192 soybean accessions of mini core collection (Kajiya-Kanegae et al., 2021) was used in this study. The SNPs dataset with minor allele frequencies of more than 5% for the 74 accessions was filtered using PLINK 2.0 (Chang et al., 2015). Association tests between the t-SNE and the SNPs dataset were conducted using a function of “association.test” in gaston package ver. 1.5.7 in R (McKenna et al., 2010). First, five principal components were included as fixed effects in the linear mixed model. A genetic relationship matrix (GRM) and ρ values of the marker-trait associations were calculated using the “GRM” function and Wald’s test, respectively. Manhattan plot and QQ plot were drawn using the “manhattan” and “qqplot.pvalues” functions, respectively. The genome-wide significant threshold was obtained based on a false discovery rate (FDR) at a 5% level using the “p.adjust” function in R (Benjamini and Hochberg, 1995).

Gene Selection

We selected all the genes in a 250 kbp region from both sides of the significant peak detected in GWAS together with annotations from the reference genome sequence Gmax_275_v2.0 from Phytozome (Phytozome v.12, **Supplementary Table S1**). About 19 genes that had either photosynthesis-related functions or higher expression in photosynthesis-related plant tissues were selected based on the Phytozome database search. The SNP data and Indels in these genes among the 74 accessions were extracted from Illumina read mapping data (Kajiya-Kanegae et al., 2021) using CLC Genomics Workbench 12 (Qiagen, Germany).

RNA Expression

On 24 August 2021, we measured the A_{sat} of the central leaflets of 14 randomly selected JMC accessions and took about 10 mg of a fresh tissue sample for gene expression analysis immediately after the A_{sat} measurement. Leaf samples were frozen in liquid nitrogen and stored at -60°C . Total RNA was extracted by TRI Reagent (Molecular Research Center, United States). First-strand cDNA was synthesized from 500 ng of total RNA by ReverTra ACE qPCR RT Master Mix with gDNA Remover (TOYOBO, Japan). Quantitative RT-PCR was performed according to the manufacturer's instructions by diluting the synthesized cDNA ten folds and using a KAPA SYBR FAST qPCR Kit (KAPA Biosystems, United States) and a ViiA7 real-time PCR system (Thermo Fisher Scientific, United States). The qRT-PCR conditions were as follows: (1) initial denaturation at 95°C for 30 s; (2) 40 cycles of denaturation at 95°C for 5 s; and (3) final annealing and elongation at 60°C for 30 s. Expression levels were standardized using a VPS-like gene (Glyma.09G196600). The primer pairs used for quantitative RT-PCR are listed in **Supplementary Table S2**.

Biochemical Parameters

Nitrogen content (N_{cont}), Chlorophyll a (*Chl a*), Chlorophyll b (*Chl b*), and Chlorophyll a + b (*Chl a + b*) were quantified at the same leaves where the gas exchange was measured. For measuring *Chl a*, *Chl b*, and *Chl a + b*; four-leaf disks of 0.5 cm diameter were taken on 30 July and dissolved in 2 ml N, N-Dimethylformamide (DMF) as described by Porra (2002). The samples were then wrapped in aluminum foil and stored at 5°C for over 24 h. Spectroscopic readings of the supernatant were taken at 646.8 nm, 663.8 nm, and 750 nm wavelengths using a U-2910 Spectrophotometer (HITACHI, Japan). Nitrogen content was measured on 30 July and 11 August. The leaf area was measured using a leaf area meter LI-3100C (LI-COR, United States). They were then oven-dried for 72 h and were used for quantifying N_{cont} using Kjeldahl digestion. We followed the method of Vickery (1946) for quantifying N_{cont} . The oven-dried leaves were crushed, and a sample of 0.2 g was dissolved and heated in 4 ml of highly concentrated sulfuric acid until the solution turned colorless. The colorless solution was then diluted to 40 ml with distilled water. We separated 20 μl aliquots of this solution into glass tubes and mixed it with 2.48 ml of distilled water, 1 ml of Indophenol A, and 1.5 ml of Indophenol B. These were then read at 635 nm wavelength using a U-1100 Spectrophotometer (HITACHI, Japan).

Morphological Parameters

Stomatal density (S_{Density}), stomatal length (S_{Length}), and stomatal width (S_{Width}) were measured on 30 July and 11 August using the same leaves used for the gas exchange measurements. We followed Tanaka and Shiraiwa (2009) for determining stomatal parameters. After taking disks for chlorophyll content analysis, we used the Suzuki Universal Method of Printing (SUMP) for printing stomatal maps. A droplet (around 10 μl)

of SUMP liquid (amyl acetate) was placed on a SUMP disk. The leaf was placed on the disk and left to air-dry for ~ 20 min. The disks were later observed at 100X (for S_{Density}) and 400X (for S_{Length} and S_{Width}) magnifications under light microscope BH-32 (OLYMPUS, Japan) with a Multi-Interface Digital Camera FLOYD (Wraymar, Japan). ImageJ was used for counting stomata and measuring their length and width. We also measured specific leaf weight (*SLW*) as an alias to leaf thickness which is the ratio of a leaf's dry weight to its area (g m^{-2}).

Statistical Tests and Graphs

Two-way ANOVA and Tukey's Honest Significant Difference (Tukey HSD) analysis were conducted for all parameters using the built-in R functions of "aov" and "TukeyHSD"; and "HSD.test" function of Agricolae package (De Mendiburu, 2015). Accessions (G) and the measurement dates (D) were used as factors. Due to field homogeneity, we considered block difference as negligible. For correlations among parameters, we used the "pearsonr" function of the Scipy library (Pauli et al., 2020) in Python. We used the Pythons matplotlib.pyplot (Hunter, 2007) library for creating figures; and Scikit learn library (Pedregosa et al., 2011) for K-means clustering.

RESULTS

Screening, Distribution, and Clustering of A_{sat}

As shown in **Figure 1**, there were great variations in A_{sat} among accessions and measurement dates. A_{sat} ranged from $10 \mu\text{mol m}^{-2} \text{s}^{-1}$ – $33 \mu\text{mol m}^{-2} \text{s}^{-1}$. Both in 2019 and 2020, the frequency of accessions with lower A_{sat} was higher at the beginning and the end of August, whereas the A_{sat} measurements conducted in mid-August had a narrow range with most accessions showing greater A_{sat} . Shapiro's test of normality revealed that the data were normally distributed.

t-SNE compressed variations of A_{sat} into two components. SNE1 explained only 22.9% variations whereas SNE2 explained 77.0% variations in A_{sat} measurements. SNE1 had a negative correlation with all the A_{sat} measurements and was particularly strong with measurements taken on 05 August 2019, and 06 August 2020. In contrast, SNE2 was positively correlated with all the A_{sat} measurements (**Table 1**). t-SNE classified accessions into four clusters with massive differences between photosynthetic capacities (**Figure 2**). The computed Silhouette coefficient was 0.76 indicating that the clusters are spherical with only minor overlapping among them.

We observed significant differences in A_{sat} among the clusters ($p < 0.001$; **Figure 3**). On average, Cluster2 had the highest A_{sat} and consisted of some of the best photosynthesizing (A_{sat}) accessions. Some of these accessions showed photosynthetic rates comparable to or higher than the elite cultivars, En in Cluster3 and Fu in Cluster1. Cluster1 and Cluster3 had cross-year variations in their A_{sat} values whereas Cluster4 was characterized by the lowest A_{sat} values in 2019 and 2020.

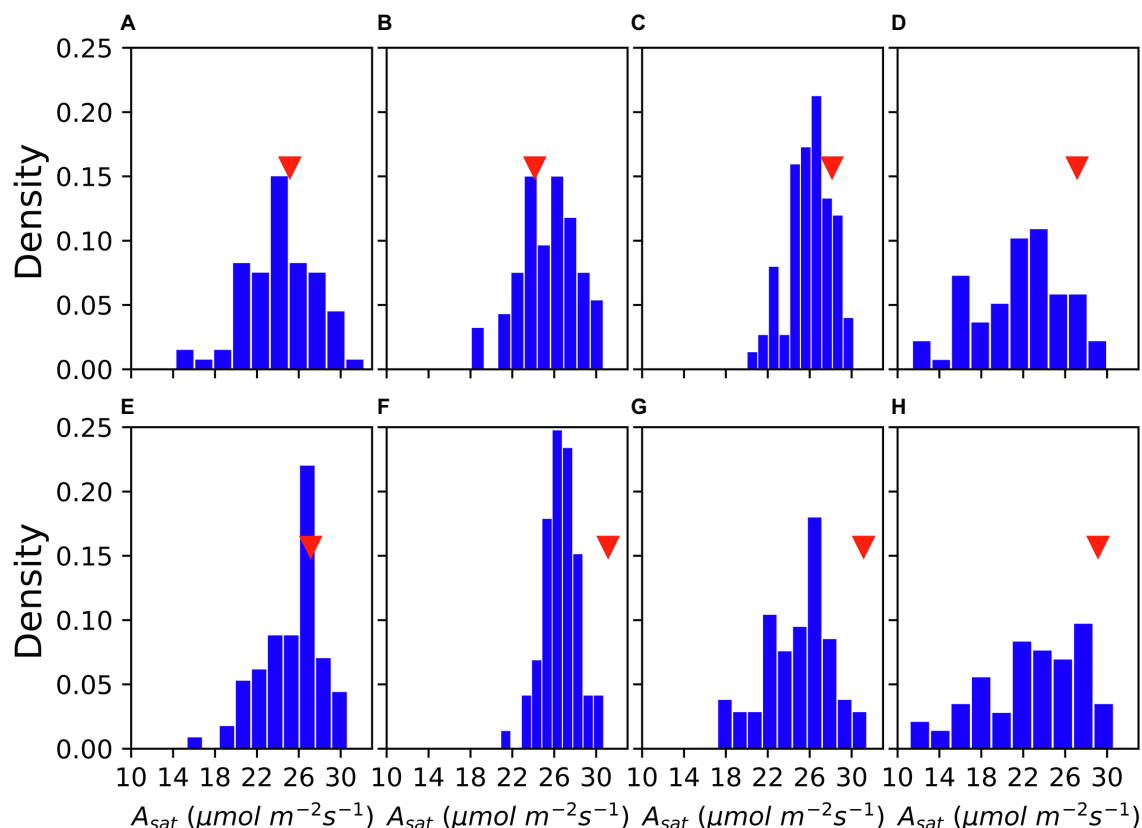


FIGURE 1 | Variation of A_{sat} in the GmJMC (JMC) accessions and its comparison with Fukuyutaka (Fu) across measurements. 5 August, 2019 (A) 13 August, 2019 (B) 18 August, 2019 (C) 6 August, 2020 (D) 14 August, 2020 (E) 21 August, 2020 (F) 27 August, 2020 (G) and 31 August, 2020 (H). Red triangles show the A_{sat} performance of Fu.

TABLE 1 | Correlation coefficient between t-SNE components and A_{sat} in each measurement date.

Date	SNE1	SNE2
5 August, 2019	-0.65	0.55
13 August, 2019	-0.13	0.52
18 August, 2019	0.13	0.57
6 August, 2020	-0.75	0.37
14 August, 2020	-0.34	0.57
21 August, 2020	-0.57	0.65
27 August, 2020	-0.13	0.76
31 August, 2020	-0.30	0.85

SNE1 and SNE2 show first and second t-SNE components, respectively.

Genetic Characterization and RNA Expression of the Candidate Genes

Based on the results of **Table 1**, SNE2 was interpreted to represent the genetic potential of A_{sat} . GWAS using SNE components revealed a region on chromosome 17 that had a strong association with variations of SNE2 (**Figure 4**). In contrast, the GWAS result of SNE1 (**Supplementary Figure S1**) was not considered for further analysis due to a lower $-\log(p)$ value and few explainable variations for A_{sat} (**Table 1**). In a 500 kbp region spanning both sides of

the significant peak detected in GWAS of SNE2, 42 genes are annotated (**Supplementary Table S1**). Among them, 19 genes were proteins that are assumed to be involved in photosynthesis-related functions. RNA expression analysis of the three candidate genes, Glyma.17G226100, Glyma.17G226700, and Glyma.17G226900, revealed a strong correlation between the expression levels of Glyma.17G226700, and A_{sat} ($r=0.72$), g_s ($r=0.69$), Transpiration or E ($r=0.62$), ETR ($r=0.56$), and leaf temperature or Tleaf ($r=-0.60$); however, there was no correlation between the expression of this gene and *intercellular CO₂ concentration* (C_i ; **Figure 5**). The correlation of Glyma.17G226100 and Glyma.17G226900 with A_{sat} is shown in **Supplementary Figure S2**.

Gas Exchange and Chlorophyll Fluorescence

Considerable variations of A_{sat} among 10 representative accessions and measurement dates were observed during our second experiment in 2021 (**Figure 6A**). A_{sat} ranged from $24.1 \text{ m}^2 \text{ s}^{-1}$ in JMC56 on 22 July to $41.5 \text{ μmol m}^{-2} \text{ s}^{-1}$ in JMC47 on 30 July. The lowest A_{sat} was observed on 22 July which then increased sharply on 30 July followed by a drop on 11 August. Among accessions, PE had the highest A_{sat} on 22 July followed by JMC47. PE and JMC47 were significantly higher than En, JMC43, and JMC56

($p < 0.05$). On 30 July, JMC47 showed the highest A_{sat} and was significantly greater than En, JMC13, JMC16, JMC49, and JMC56 ($p < 0.01$). It was also the best performing accession on 11 August, but the difference was not significant among accessions. The g_s measurements showed a similar pattern to A_{sat} , but with greater magnitudes (Figure 6B). It ranged from 0.46 in JMC43 on 22

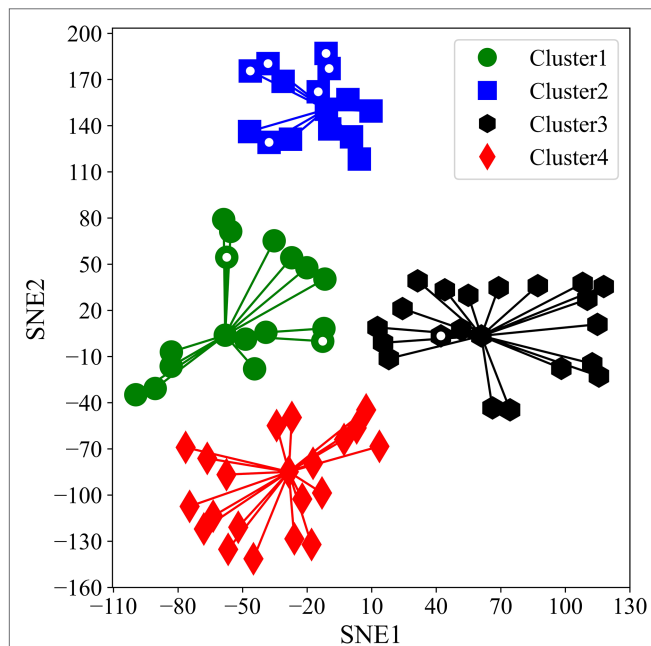


FIGURE 2 | Two-dimensional scatter plot of four different K-means clusters alongside SNE1 and SNE2. White dots inside markers indicate representative accessions we used for our second experiment.

July to 1.69 mol m⁻² s⁻¹ in JMC47 on 30 July. The g_s values were exceptionally low on 22 July followed by a sharp increase on 30 July and a drop on 11 August. PE and JMC47 had the highest g_s values in all measurements. Seasonal change in C_i was also comparable to A_{sat} and g_s . The lowest rates of C_i were recorded on 22 July, followed by an increase on 30 July and a drop on 11 August (Figure 6C). Among accessions, JMC43 had the lowest C_i of 273.3 μmol mol⁻¹ on 22 July, whereas JMC56 had a C_i of 313.6 μmol mol⁻¹ on 22 July. We also calculated the apparent mesophyll activity (A/C_i) and found that its variations were strongly correlated with variations in A_{sat} (Supplementary Table S3). JMC47 had the highest A/C_i values in all measurements whereas En and JMC56 had the lowest A/C_i on 22 July.

Chlorophyll fluorescence parameters exhibited a similar pattern to gas exchange parameters. On average, ETR was significantly lower on 22 July than 30 July and 11 August ($p < 0.05$) but the values were not significantly different between 30 July and 11 August (Figure 6D). ETR ranged from 212.1 μmol m⁻² s⁻¹ in JMC56 to 276.2 μmol m⁻² s⁻¹ in JMC47. Fv'/Fm' rates were also the lowest on 22 July followed by an increase on 30 July and a drop on 11 August (Figure 6E). JMC47 exhibited the highest Fv'/Fm' among all the measurements. In contrast, $NPQ(t)$ exhibited a reverse pattern of gas exchange and fluorescence parameters (Figure 6F). Its mean values were 1.65, 1.12, and 1.33 on 22 July, 30 July, and 11 August, respectively, and the difference was significant ($p < 0.01$). It ranged from 0.91 in JMC47 on 30 July to 1.8 in En on 22 July. JMC47 had significantly low $NPQ(t)$ values on 22 July and 30 July among all the accessions ($p < 0.01$).

A_{sat} was strongly correlated with g_s in all the measurements; however, the strength of correlation was weaker on 11 August (Supplementary Figure S4). The positive correlation between

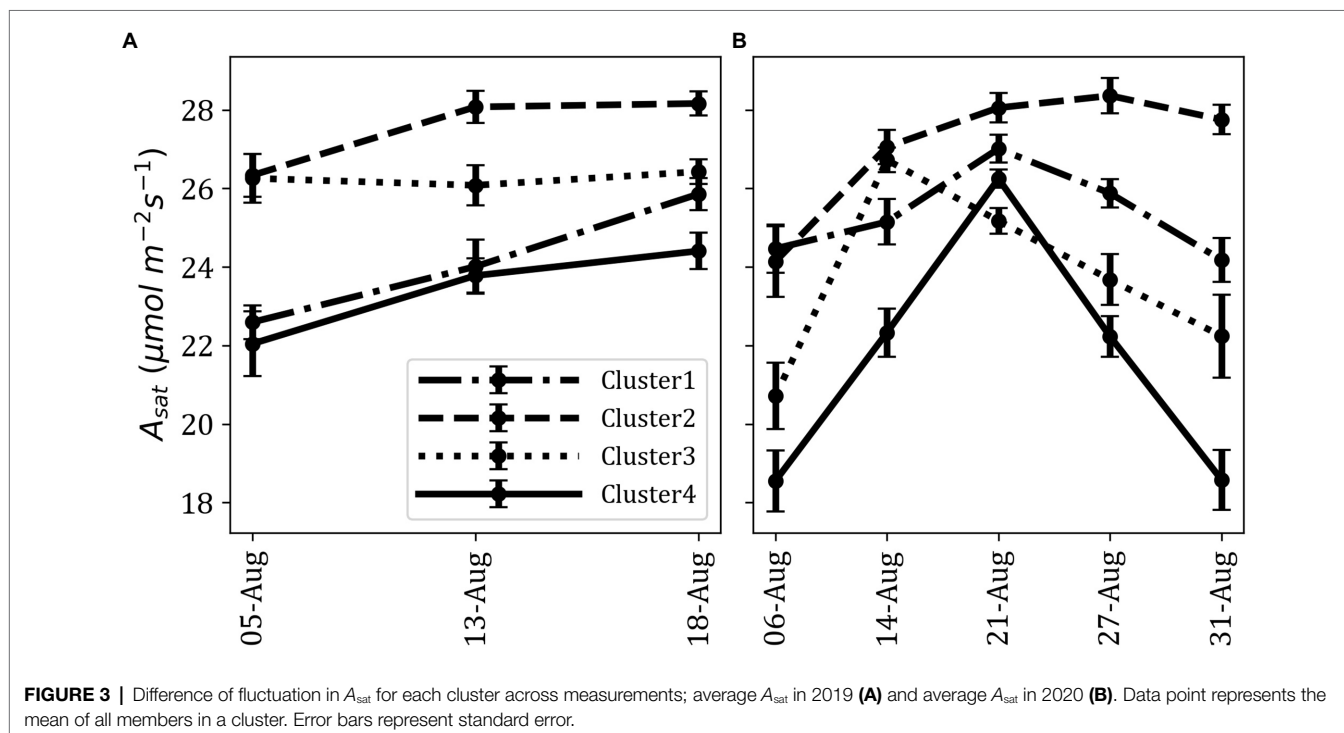


FIGURE 3 | Difference of fluctuation in A_{sat} for each cluster across measurements; average A_{sat} in 2019 (A) and average A_{sat} in 2020 (B). Data point represents the mean of all members in a cluster. Error bars represent standard error.

A_{sat} and C_i was significant on 22 July ($p < 0.01$) whereas it became negatively correlated in the last two measurements. Interestingly, the positive correlation between C_i and g_s was constantly observed across all the measurements. ETR , F_v/F_m' , and $PhiPS2$ had strong correlations with A_{sat} . On contrary, $NPQ(t)$ was negatively correlated with A_{sat} and other Chlorophyll fluorescence parameters.

Chlorophyll and Nitrogen Content

Variations in *Chl a*, *Chl b*, and *Chl a+b* contents were observed among accessions. En, JMC49, and JMC56 are distinguished with lower *Chl a*, *Chl b*, and *Chl a+b* contents. *Chl a* was significantly ($p < 0.05$) higher in PE than in En and JMC56, and *Chl b* was significantly higher in PE than En, JMC13, and JMC56 ($p < 0.05$; **Supplementary Figure S3A**). JMC47 had the highest *Chl a*, *Chl b*, and *Chl a+b* contents among the JMC accessions and was comparable with that of PE. We observed consistency of N_{cont} variation among accessions between 30 July and 11 August except for JMC16 and JMC56 (**Supplementary Figure S3B**). The lowest N_{cont} of 1.25 gm^{-2} was seen in JMC37 and JMC43 whereas JMC56 had a higher N_{cont} value of 1.92 gm^{-2} on 30 July than that of most accessions. We did not observe any significant difference in N_{cont} values on 11 August. On average, N_{cont} was higher on 30 July than on 11 August.

The correlation between *Chl a*, *Chl b*, and *Chl a+b* and A_{sat} was significant ($p < 0.01$; **Supplementary Figure S4**). However, we did not observe a significant correlation between chlorophyll contents and chlorophyll fluorescence parameters.

Stomatal Attributes and Specific Leaf Weight

JMC47 had the highest S_{Density} among the JMC accessions on 30 July (**Figure 7A**). PE and JMC47 had significantly more stomata than Fu, JMC13, JMC37, and JMC43 ($p < 0.05$). The S_{Density} difference among accessions was not clear on 11 August. Variations of S_{Length} and S_{Width} (**Figure 7B**) were also observed among accessions and between measurements, but the difference was only significant between the measurement dates ($p < 0.05$) and not among the accessions. S_{Length} was twofold longer than S_{Width} for all the accessions.

En and JMC47 showed distinguishably higher specific leaf weight (SLW) on 30 Jul, however, there was no consistency between the values of 30 July and 11 August (**Figure 7C**). JMC13 and JMC49 had smaller SLW values than other accessions on 30 July; however, on 11 August, PE and JMC56 had the lowest SLW values. We observed accessions (G) to measurement date (D) interaction at $p < 0.05$, but the average SLW values were neither significant among accessions nor between the two measurement dates.

A positive correlation between gas exchange parameter g_s and S_{Density} was observed (**Supplementary Figure S4**) but it was only significant on 11 August ($p < 0.05$). A strong negative correlation existed between S_{Density} and S_{Length} whereas the correlation between S_{Density} and S_{Width} was not significant. S_{Length} was also negatively correlated with A_{sat} , C_i , and g_s . SLW was not correlated with A_{sat} in pooled data, nor was it with individual measurements.

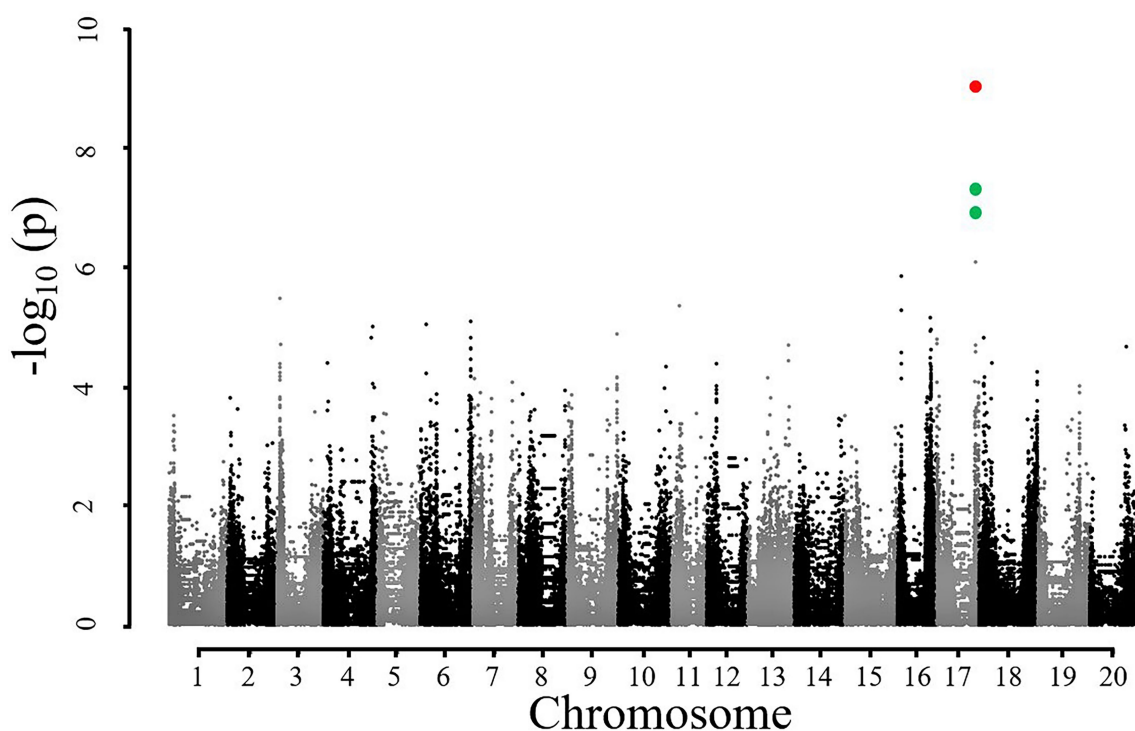


FIGURE 4 | Manhattan plot for showing significant association of SNPs with SNE2. Red and green dots on chromosome 17 represent significant and suggestive associations where $-\log_{10}(p)$ value is less than 8 and 5, respectively.

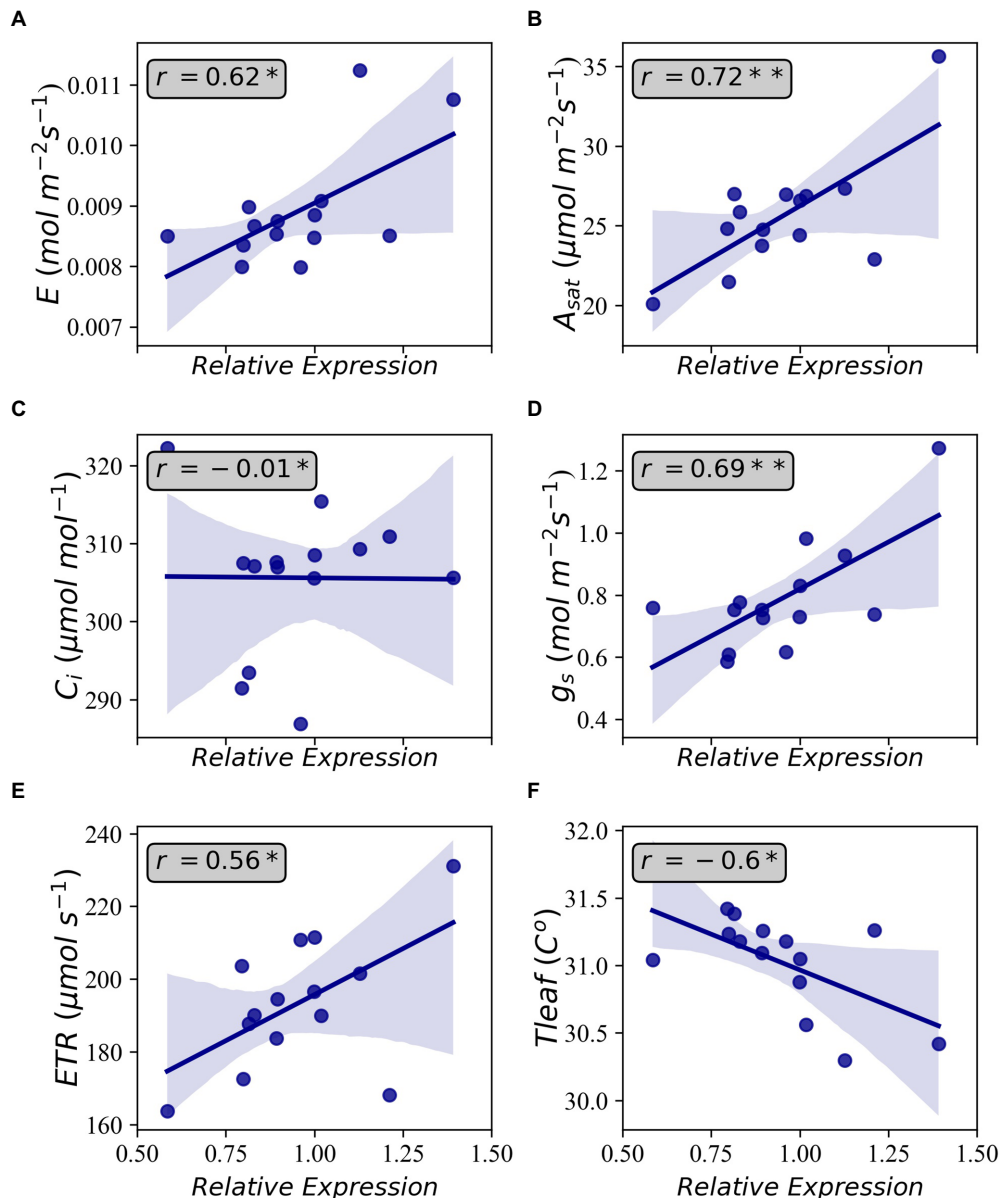


FIGURE 5 | Correlations between relative gene expression of Glyma.17G226700 and Transpiration (A), light-saturated photosynthesis (B), intercellular CO_2 concentration (C), stomatal conductance (D), electron transfer rate (E), and leaf temperature (F). Shaded areas indicate 95 percent confidence intervals. Annotations inside the plots indicate Pearson's correlation coefficient. Each data point is the average of four observations. * $p < 0.05$ and ** $p < 0.01$.

DISCUSSION

The culmination of conventional yield improving parameters has widened the margin between food demand and crop yield (Zhu et al., 2008, 2010), leaving the potential yield productivity to be bridged by the manipulation of photosynthetic processes in plants (Long et al., 2006; Kromdijk et al., 2016). Although exploiting natural variations of photosynthetic apparatus is considered promising for increasing crop productivity (Heckmann et al., 2017), unexplored germplasm is believed to be a major bottleneck (Kromdijk and Long, 2016) toward achieving it.

We studied leaf photosynthetic capacity in JMC to (1) harness natural genotypic variations of photosynthetic capacity; (2) study the underlying genetic architecture of light-saturated photosynthesis; (3) identify physiological traits underpinning photosynthetic variations. This study is the first to assess JMC for photosynthetic studies on this scale.

t-SNE and K-Means Clustering

As photosynthesis is a physiological process that could be affected by various environmental factors, we considered it was of importance to conduct multiple and repeated measurements

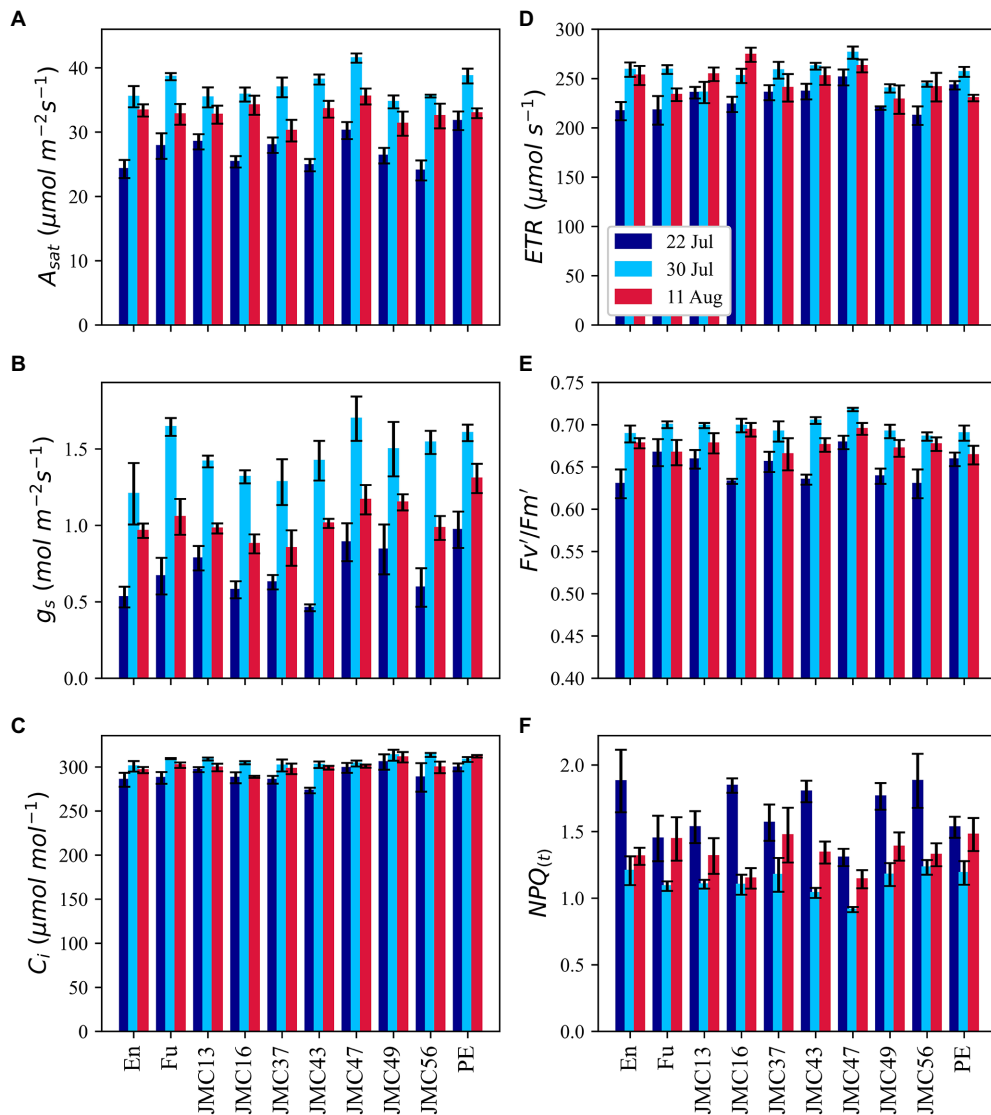
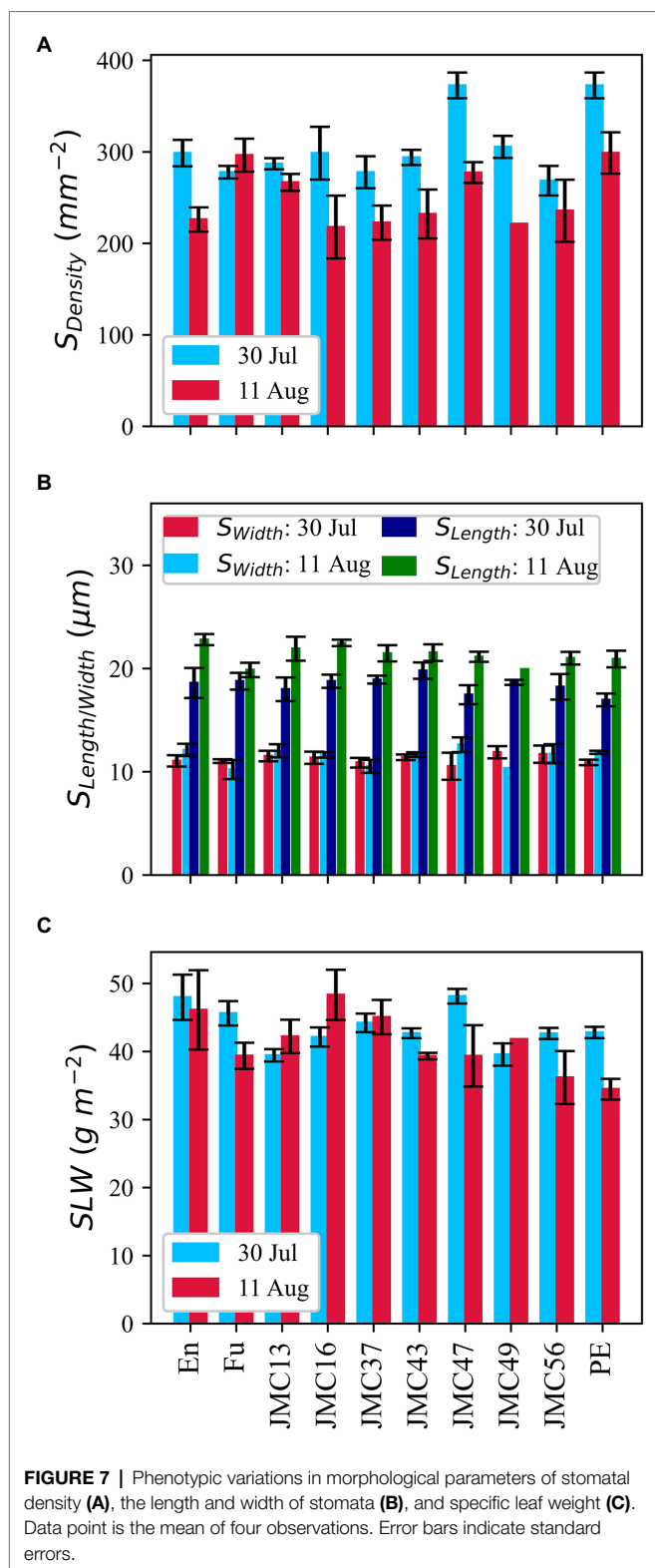


FIGURE 6 | Phenotypic variations in gas exchange and chlorophyll fluorescence parameters; light-saturated photosynthetic rate (A), stomatal conductance (B), intercellular CO_2 concentration (C), electron transport rate (D), the light utilization efficiency of PSII (E) and alias to non-photochemical quenching (F). Data point is the mean of four observations. Error bars indicate standard error.

and then explain total photosynthetic variations over a time course *via* dimensionality reduction. The use of MIC-100 enabled us to measure A_{sat} repeatedly in a stunningly high number of plants which is not possible using conventional gas exchange measurement systems. The method of aggregation and interpretation of a large amount of data through large-scale phenotyping will be important to solve the bottleneck of harnessing limited genetic variations in photosynthetic research. Dimensionality reduction combined with GWAS has previously been used by Yano et al. (2019). The use of t-SNE has important implications. t-SNE is one of the best dimensionality reduction algorithms that work well for both linear and non-linear data. It is also a very robust method for visualizing high dimensional data in two- or three-dimensional

panels which we considered using for our analysis. Our t-SNE analysis revealed some suitable photosynthetic characteristics of the JMC. K-means clustering classified the studied accessions into four distinct clusters with varying A_{sat} values (Figures 2, 3). The correlation between actual data and t-SNE components indicates that SNE2 represents the genetic potential of photosynthetic capacity, whereas SNE1 seems to explain seasonal variations in A_{sat} because of their high correlation with measurements at the beginning of August in both years (Table 1). The difference in photosynthetic performance among these clusters could be attributed to some geographical or agro-morphological basis, in addition to the unexplored limitations that lie within photosynthetic machinery. Among four clusters, Cluster2 is of particular interest as its members showed their



high photosynthetic rates equivalent to or higher than the photosynthetically elite cultivars, Fu in Cluster1 and En in Cluster3 (Figure 2). Higher A_{sat} values among the JMC accessions serve as a source of evidence for the claim that harnessing

natural variations of photosynthesis is an ideal route toward removing bottlenecks in this process and improving seed yield (Heckmann et al., 2017).

Genetic Architecture of A_{sat}

The use of SNE2 in GWAS has two important implications. Firstly, SNE2 explains a great portion of variations in A_{sat} values, and secondly, it had a strong positive correlation with all the measurement dates (Table 1). GWAS using SNE2 identified a genomic region at chromosome 17 (Figure 4) which had a strong correlation with SNE2. The accession numbers used in the present study was relatively small for GWAS and the detected peak should be interpreted carefully. However, the repeated measurement of A_{sat} in 8 environments across 2 years and subsequent dimensionality reduction enabled us to detect the possible genomic region controlling the photosynthetic capacity.

One of the genes near the identified SNP is Glyma.17G226700, encoding G protein alpha subunit 1 (GPA1) which binds to abscisic acid (ABA). GPA1 mutants have been reported to have increased transpiration efficiency by minimizing g_s and stomatal density (Nilson and Assmann, 2010). The loss-of-function mutation of GPA1 led to a decrease in stomatal density in the lower epidermis of *Arabidopsis* cotyledons suggesting that GPA1 positively regulates stomatal density in response to environmental and developmental cues (Zhang et al., 2008). GPA1 has also been reported to be dynamically regulating chloroplast development and response mechanism to both intracellular and intercellular signals (Zhang et al., 2009). In a previous study on soybean, Glyma.17G226700 was also detected through GWAS of photosynthetic traits concerning phosphorus efficiency (Lü et al., 2018). The positive relationship between the gene expression level of Glyma.17G226700 and A_{sat} (Figure 5B) in this study suggests that GPA1 is an important factor in determining the natural variation of photosynthetic capacity in soybean. Nevertheless, further analysis on elucidating the functions of this gene in regulating photosynthesis is needed to utilize genetic resources in future breeding programs. Other recent studies focusing on the genetic architecture of photosynthetic capacity have also reported several loci. None of them, however, overlapped with our study. Lopez et al. (2019) identified loci related to photosynthetic capacity on chromosomes 3 and 15 in the NAM population of the US soybean. Wang et al. (2020) also reported loci affecting photosynthetic rate on chromosomes 6, 16, 18, and 19 in Chinese soybean germplasm. Yang et al. (2022) identified some loci controlling chlorophyll fluorescence on chromosomes 2, 4, 7, 8, 14, 15, 16, 18, 19, and 20 in Chinese soybean germplasm. The detection of various genomic regions between previous reports and our study suggests the complex genetic regulation and environmental interaction of soybean leaf photosynthesis.

A_{sat} and Physiological Mechanism Underpinning Photosynthetic Capacity

Photosynthetic capacity of the JMC accessions, in general, was highly dependent on CO_2 diffusion, mainly due to stomatal

limitations (**Figure 6**). Low g_s on 22 July could be attributed to the early vegetative stages of the accessions. The implication of g_s on increasing photosynthetic capacity is well-established (Cornish et al., 1991; Fay et al., 1995; Taylor and Long, 2017; Yamori et al., 2020). The correlation between g_s and S_{Density} (**Supplementary Figures S4B,C**) and the higher g_s and S_{Density} in JMC47 and PE indicates that higher g_s in JMC47 and PE was due to their leaves allocating more surface to stomata. Our results are consistent with the findings of Franks et al. (2009) and Tanaka et al. (2010) also reported a strong correlation between potential stomatal conductance and S_{Density} and argued that it could be underpinning the physiological potential for greater productivity in the U.S. soybean cultivars. In this study, JMC accessions with higher A_{sat} had higher g_s and higher g_s was usually accompanied by a higher S_{Density} (**Figure 7**). This indicates that Japanese soybean may show greater photosynthetic capacity if their stomatal g_s is manipulated.

The best performance of JMC47 throughout vegetative and reproductive growth stages appears to be due to enhanced light utilization in addition to higher g_s . This accession displayed the highest ETR , Fv'/Fm' , Φ_{PSII} , and the lowest $NPQ(t)$ values (**Figure 6**) indicating that it utilizes sunlight much more efficiently than the elite cultivar of Fu, En, and PE. The continuous superiority of ETR and Fv'/Fm' in JMC47 across measurement dates is consistent with findings that the application of 24-epibrassinolide increased ETR and the net-photosynthetic rates under water-deficient conditions and suggested that plants with higher ETR are more efficient in photosynthesis (Pereira et al., 2019). Sheng et al. (2008) studied the effect of *Arbuscular mycorrhiza* (AM) on photosynthesis and water status in maize and found that enhancement in photosynthetic capacity was due to enhanced Fv'/Fm' , Fv/Fm , and Φ_{PSII} , and lower non-photochemical quenching (NPQ). Although we did not use any treatment, the natural superiority of JMC47 in showing greater Fv'/Fm' could be utilized as a potential resource in enhancing the light use efficiency of the photosynthetic process.

The relationship between chlorophyll content and photosynthesis has been studied intensively (Buttery and Buzzell, 1977; Garty et al., 2001; Li et al., 2006; Holly et al., 2017) and a strong positive correlation between leaf chlorophyll content and photosynthesis has been reported. A positive correlation (**Supplementary Figure S4B**) between chlorophyll content ($Chl\ a$, $Chl\ b$, $Chl\ a+b$) and A_{sat} falls in agreement with previous findings. Higher A_{sat} in JMC47 could also be related to its increased chlorophyll content; however, the higher chlorophyll content in PE but lower A_{sat} compared to that of JMC47 may be due to other yet unidentified factors. N_{cont} and SLW are correlated with A_{sat} (Allison et al., 1997; Boussadia et al., 2010; Sakoda et al., 2016; Shamim et al., 2021). In our study, however, we did not observe any correlation. These results indicate that JMC47 had both stomatal and light utilization advantages over other accessions that led to its highest photosynthetic capacity.

This is the first study that reports detailed information about the photosynthetic capacity of Japanese soybean accessions together with physiological mechanism and genetic architecture

underpinning photosynthesis. Here, we classified Japanese soybean germplasm into four clusters with Cluster2 comprising highly photosynthesizing accessions. GWAS revealed a locus on chromosome 17 and further genomic studies suggested the involvement of a gene, Glyma.17G226700, whose expression had a strong positive correlation with A_{sat} . Experiment on the ten representative accessions revealed that there were stomatal and non-stomatal limitations on A_{sat} with major limitations due to CO_2 diffusion (lower S_{Density}). Among all accessions, JMC47 had the highest A_{sat} which could be attributed to its high g_s , S_{Density} , chlorophyll content, and chlorophyll fluorescence parameters. These results provide a piece of solid evidence to suggest that exploiting the genetic variations of plant germplasms is an ideal route to unlocking the resource use efficiency in photosynthesis.

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

MS and YT designed the experiment. MS conducted the experiment, data analysis, and interpretation. GWAS was conducted by AK. HY analyzed RNA expression. The manuscript was written by MS with a contribution to the methodology of GWAS, Gene selection, and RNA expression from AK. AK and YT reviewed the manuscript and acquired funding. YT, AK, and TS coordinated and supervised the research. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by a grant from the Ministry of Agriculture, Forestry, and Fisheries of Japan (Smart-Breeding System for Innovative Agriculture, BAC1003) and Ministry of Education, Culture, Sport, Science, and Technology supported MS.

ACKNOWLEDGMENTS

We express our gratitude to the Ministry of Agriculture, Forestry, and Fisheries; and the Ministry of Education, Culture, Sport, Science, and Technology of the Japanese government for supporting the research and the student.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.910527/full#supplementary-material>

REFERENCES

- Allison, J. C. S., Williams, H. T., and Pammenter, N. W. (1997). Effect of specific leaf nitrogen content on photosynthesis of sugarcane. *Ann. Appl. Biol.* 131, 339–350. doi: 10.1111/j.1744-7348.1997.tb05160.x
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Boussadia, O., Steppe, K., Zgallai, H., Ben El Hadj, S., Braham, M., Lemeur, R., et al. (2010). Effects of nitrogen deficiency on leaf photosynthesis, carbohydrate status and biomass production in two olive cultivars “Meski” and “Koroneiki”. *Sci. Hortic.* 123, 336–342. doi: 10.1016/j.scienta.2009.09.023
- Buttery, B. R., and Buzzell, R. L. (1977). The relationship between chlorophyll content and rate of photosynthesis in soybeans. *Can. J. Plant Sci.* 61, 190–197. doi: 10.4141/cjps81-029
- Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S., and Lee, J. (2015). Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* 4:7. doi: 10.1186/s13742-015-0047-8
- Christensen, J. H., and Hewitson, B. (2007). “Regional climate projections,” in *Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. eds. S. Solomon, M. Manning, M. Marquis and D. Qin (Cambridge: Cambridge University press).
- Cornish, K., Radin, J. W., Turcotte, E. L., Lu, Z., and Zeiger, E. (1991). Enhanced photosynthesis and stomatal conductance of pima cotton (*Gossypium barbadense* L.) bred for increased yield. *Plant Physiol.* 97, 484–489. doi: 10.1104/pp.97.2.484
- De Mendiburu, F. D. (2015). *Agricolae: statistical procedures for agricultural research. R Package Version 1*, 2–3.
- Fay, P. A., Knapp Eay, A. K., and Fay, P. A. (1995). Stomatal and photosynthetic responses to shade in sorghum, soybean and eastern gamagrass. *Physiol. Plant.* 94, 613–620. doi: 10.1111/j.1399-3054.1995.tb00975.x
- Franks, P. J., Drake, P. L., and Beerling, D. J. (2009). Plasticity in maximum stomatal conductance constrained by negative correlation between stomatal size and density: An analysis using *Eucalyptus globulus*. *Plant Cell Environ.* 32, 1737–1748. doi: 10.1111/j.1365-3040.2009.02031.x
- Gabriel, A. S., Ninomiya, K., and Uneyama, H. (2018). The role of the Japanese traditional diet in healthy and sustainable dietary patterns around the world. *Nutrients* 10:173. doi: 10.3390/nu10020173
- Garty, J., Tamir, O., Hassid, I., Eshel, A., Cohen, Y., Karnieli, A., et al. (2001). Photosynthesis, chlorophyll integrity, and spectral reflectance in lichens exposed to air pollution. *J. Environ. Qual.* 30, 884–893. doi: 10.1007/978-81-322-2181-4_3
- Hay, W. T., Bihmidine, S., Mutlu, N., Le Hoang, K., Awada, T., Weeks, D. P., et al. (2017). Enhancing soybean photosynthetic CO₂ assimilation using a cyanobacterial membrane protein, *ictB*. *J. Plant Physiol.* 212, 58–68. doi: 10.1016/j.jplph.2017.02.003
- Heckmann, D., Schlüter, U., and Weber, A. P. M. (2017). Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra. *Mol. Plant* 10, 878–890. doi: 10.1016/j.molp.2017.04.009
- Holly, C., Jing, M. C., Xiangzhong, L., Paul, B., Bin, C., and Ralf, M. S. (2017). Leaf chlorophyll content as a proxy for leaf photosynthetic capacity. *Glob. Change Biol.* 23, 3513–3524. doi: 10.1111/gcb.13599
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Comp. Sci. Eng.* 9, 90–95. doi: 10.1109/MCSE.2007.55
- Janssen, P. J. D., Lambrev, M. D., Plummer, N., Bartolucci, C., Antonacci, A., Buonassera, K., et al. (2014). Photosynthesis at the forefront of a sustainable life. *Front. Chem.* 2:36. doi: 10.3389/fchem.2014.00036
- Kaga, A., Shimizu, T., Watanabe, S., Tsubokura, Y., Katayose, Y., Harada, K., et al. (2012). Evaluation of soybean germplasm conserved in NIAS genebank and development of mini core collections. *Breed. Sci.* 61, 566–592. doi: 10.1270/jsbbs.61.566
- Kajiya-Kanegae, H., Nagasaki, H., Kaga, A., Hirano, K., Ogiso-Tanaka, E., Matsuoka, M., et al. (2021). Whole-genome sequence diversity and association analysis of 198 soybean accessions in mini-core collections. *DNA Res.* 28:dsaa032. doi: 10.1093/dnares/dsaa032
- Kawasaki, Y., Tanaka, Y., Katsura, K., Purcell, L. C., and Shiraiwa, T. (2016). Plant production science yield and dry matter productivity of Japanese and US soybean cultivars. *Plant Prod. Sci.* 19, 257–266. doi: 10.1080/1343943X.2015.1133235
- Khush, G. S. (2001). Green revolution: the way forward. *Nat. Rev. Genet.* 2, 815–822. doi: 10.1038/35093585
- Kromdijk, J., Glowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., et al. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* 354, 857–861. doi: 10.1126/science.aai8878
- Kromdijk, J., and Long, S. P. (2016). One crop breeding cycle from starvation? How engineering crop photosynthesis for rising CO₂ and temperature could be one important route to alleviation. *Proc. Royal Soc. B* 283:20152578. doi: 10.1098/rspb.2015.2578
- Kusumi, K., Hirotsuka, S., Kumamura, T., and Iba, K. (2012). Increased leaf photosynthesis caused by elevated stomatal conductance in a rice mutant deficient in SLAC1, a guard cell anion channel protein. *J. Exp. Bot.* 63, 5635–5644. doi: 10.1093/jxb/err313
- Lawlor, D. W., Paul, M. J., and Foyer, C. H. (1999). Genetic manipulation of photosynthesis. *J. Exp. Bot.* 50, 4–12.
- Li, R.-H., Guo, P.-P., Baumz, M., Grand, S., and Ceccarelli, S. (2006). Evaluation of chlorophyll content and fluorescence parameters as indicators of drought tolerance in barley. *Agri. Sci. China* 5, 751–757. doi: 10.1016/S1671-2927(06)60120-x
- LI-COR Biosciences (2021). LI-6800 portable photosynthesis system: advancements in gas exchange and fluorescence research. Available at: <https://www.licor.com/documents/02057bkvtf900ijr91z9gcwiyzwzic46> (Accessed March 22, 2022).
- Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330. doi: 10.1111/j.1365-3040.2005.01493
- Lopez, M. A., Xavier, A., and Rainey, K. M. (2019). Phenotypic variation and genetic architecture for photosynthesis and water use efficiency in soybean (*Glycine max*. Merr). *Front. Plant Sci.* 10:680. doi: 10.3389/fpls.2019.00680
- Lü, H., Yang, Y., Li, H., Liu, Q., Zhang, J., Yin, J., et al. (2018). Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. *Front. Plant Sci.* 9:1226. doi: 10.3389/fpls.2018.01226
- Maurino, V. G., and Weber, A. P. M. (2013). Engineering photosynthesis in plants and synthetic microorganisms. *J. Exp. Bot.* 64, 743–751. doi: 10.1093/jxb/ers263
- 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
- Nilson, S. E., and Assmann, S. M. (2010). The α -subunit of the Arabidopsis heterotrimeric G protein, GPA1, is a regulator of transpiration efficiency. *Plant Physiol.* 152, 2067–2077. doi: 10.1104/PP.109.148262
- Parry, M. A. J., Adnralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-silva, E., Alonso, H., et al. (2012). Rubisco activity and regulation as targets for crop improvement. *J. Exp. Bot.* 64, 717–730. doi: 10.1093/jxb/ers336
- Pauli, V., Ralf, G., Travis, E. O., Matt, H., Tylor, R., David, C., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272. doi: 10.1038/s41592-019-0686-2
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 12, 2825–2830. doi: 10.5555/1953048.2078195
- Pereira, Y., Rodrigues, W., Lima, E., Santos, L., Silva, M., and Lobato, A. (2019). Brassinosteroids increase electron transport and photosynthesis in soybean plants under water deficit. *Photosynthetica* 57, 181–191. doi: 10.32615/ps.2019.029
- Porra, R. J. (2002). The chequered history of the development and use of simultaneous equations for the accurate determination of chlorophylls and b. *Photosyn. Res.* 73, 149–156. doi: 10.1023/A:1020470224740
- Ray, D. K., Mueller, N. D., West, P. C., and Foley, J. A. (2012). Yield trends are insufficient to double global crop production by 2050. *PLoS One* 8:e66428. doi: 10.1371/journal.pone.0066428
- Sakoda, K., Tanaka, Y., Long, S. P., and Shiraiwa, T. (2016). Genetic and physiological diversity in the leaf photosynthetic capacity of soybean. *Crop Sci.* 56, 2731–2741. doi: 10.2135/cropsci2016.02.0122
- Shamim, M. J., Tanaka, Y., Sakoda, K., Shiraiwa, T., and Nelson, R. L. (2021). Physiological analysis of leaf photosynthesis of backcross-derived progenies from soybean (*Glycine max* (L.) Merrill) and *G. tomentella* Hayata. *Plant Prod. Sci.* 24, 109–117. doi: 10.1080/1343943X.2020.1807369

- Sheng, M., Tang, M., Chen, H., Yang, B., Zhang, F., and Huang, Y. (2008). Influence of arbuscular mycorrhizae on photosynthesis and water status of maize plants under salt stress. *Mycorrhiza* 18, 287–296. doi: 10.1007/s00572-008-0180-7
- Simkin, A. J., McAusland, L., Headland, L. R., Lawson, T., and Raines, C. A. (2015). Multigene manipulation of photosynthetic carbon assimilation increases CO₂ fixation and biomass yield in tobacco. *J. Exp. Bot.* 66, 4075–4090. doi: 10.1093/jxb/erv204
- Stapper, M., and Fischer, R. A. (1990). Genotype, sowing date, and plant spacing influence on high-yielding irrigated wheat in southern New South Wales. 3. Potential yields and optimum flowering dates. *Aust. J. Agric. Res.* 41, 1043–1056. doi: 10.1071/AR9900997
- Takahara, K., Kasajima, I., Takahashi, H., Hashida, S. N., Itami, T., Onodera, H., et al. (2010). Metabolome and photochemical analysis of rice plants overexpressing Arabidopsis NAD kinase. *Gene Plant Physiol.* 152, 1863–1873. doi: 10.1104/pp.110.153098
- Tanaka, Y., Fujii, K., and Shiraiwa, T. (2010). Variability of leaf morphology and stomatal conductance in soybean [*Glycine max* (L.) Merr.] cultivars. *Crop Sci.* 50, 2525–2532. doi: 10.2135/cropsci2010.02.0058
- Tanaka, Y., and Shiraiwa, T. (2009). Stem growth habit affects leaf morphology and gas exchange traits in soybean. *Ann. Bot.* 104, 1293–1299. doi: 10.1093/aob/mcp240
- Tanaka, Y., Sugano, S. S., Shimada, T., and Hara-Nishimura, I. (2013). Enhancement of leaf photosynthetic capacity through increased stomatal density in *Arabidopsis*. *New Phytol.* 198, 757–764. doi: 10.1111/nph.12186
- Tanaka, Y., Taniyoshi, K., Imamura, A., Mukai, R., Sukemura, S., Sakoda, K., et al. (2021). MIC-100, a new system for high-throughput phenotyping of instantaneous leaf photosynthetic rate in the field. *Funct. Plant Biol.* 49, 496–504. doi: 10.1071/FP21029
- Taylor, S. H., and Long, S. P. (2017). Slow induction of photosynthesis on the shade to sun transitions in wheat may cost at least 21% of productivity. *Phil. Trans. R. Soc. B* 372:20160543. doi: 10.1098/rstb.2016.0543
- Tietz, S., Hall, C. C., Cruz, J. A., and Kramer, D. M. (2017). NPQ(T): a chlorophyll fluorescence parameter for rapid estimation and imaging of non-photochemical quenching of excitons in photosystem-II-associated antenna complexes. *Plant Cell Environ.* 40, 1243–1255. doi: 10.1111/pce.12924
- Tilman, D., Balzer, C., Hill, J., and Befort, B. L. (2011). Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20260–20264. doi: 10.1073/PNAS.1116437108
- Upadhyaya, H. D. (2015). Establishing core collections for enhanced use of germplasm in crop improvement. *J. Crop Breed. Gen.* 1, 1–12.
- Vickery, H. B. (1946). The early years of the Kjeldahl method to determine nitrogen. *Yale J. Biol. Med.* 18, 473–514.
- Vieira, A. J. D., De Oliveira, D. A., Soares, T. C. B., Schuster, I., Piovesan, N. D., Martínez, C. A., et al. (2006). Use of the QTL approach to the study of soybean trait relationships in two populations of recombinant inbred lines at the F7 and F8 generations. *Braz. J. Plant Physiol.* 18, 281–290. doi: 10.1590/S1677-04202006000200004
- Wang, L., Yang, Y., Zhang, S., Che, Z., Yuan, W., and Yu, D. (2020). GWAS reveals two novel loci for photosynthesis-related traits in soybean. *Mol. Genet. Genomics* 295, 705–716. doi: 10.1007/S00438-020-01661-1
- Xiao, L., Liu, X., Lu, W., Chen, P., Quan, M., Si, J., et al. (2020). Genetic dissection of the gene coexpression network underlying photosynthesis in *Populus*. *Plan. Biotechnol.* 18, 1015–1026. doi: 10.1111/pbi.13270
- Yamori, W., Kusumi, K., Iba, K., and Terashima, I. (2020). Increased stomatal conductance induces rapid changes to photosynthetic rate in response to naturally fluctuating light conditions in rice. *Plant Cell Environ.* 43, 1230–1240. doi: 10.1111/pce.13725
- Yang, Y., Wang, L., Che, Z., Wang, R., Cui, R., Xu, H., et al. (2022). Novel target sites for soybean yield enhancement by photosynthesis. *Plant Physiol.* 268:153580. doi: 10.1016/J.JPLPH.2021.153580
- Yano, K., Morinaka, Y., Wang, F., Huang, P., Takehara, S., Hirai, T., et al. (2019). GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proc. Natl. Acad. Sci. U. S. A.* 116, 21262–21267. doi: 10.1073/pnas.1904964116
- Zhang, L., Hu, G., Cheng, Y., and Huang, J. (2008). Heterotrimeric G protein α and β subunits antagonistically modulate stomatal density in *Arabidopsis thaliana*. *Dev. Bio.* 324, 68–75. doi: 10.1016/j.ydbio.2008.09.008
- Zhang, L., Wei, Q., Wu, W., Cheng, Y., Hu, G., Hu, F., et al. (2009). Activation of the heterotrimeric G protein α -subunit GPA1 suppresses the ftsh-mediated inhibition of chloroplast development in *Arabidopsis*. *Plant J.* 58, 1041–1053. doi: 10.1111/j.1365-313X.2009.03843.x
- Zhu, X. G., Long, S. P., and Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Curr. Opin. Biotechnol.* 19, 153–159. doi: 10.1016/j.copbio.2008.02.004
- Zhu, X., Long, S. P., and Ort, D. R. (2010). Improving photosynthetic efficiency for greater yield. *Annu. Rev. Plant Biol.* 61, 235–261. doi: 10.1146/annurev-arplant-042809-112206

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 Shamim, Kaga, Tanaka, Yamatani and Shiraiwa. 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.



Improving Ultra-Low Temperature Preservation Technologies of Soybean Pollen for Off-Season and Off-Site Hybridization

Hongchang Jia^{1,2†}, Xin Liang^{1†}, Lixin Zhang^{1†}, Jinmei Zhang^{1†}, Enoch Sapey^{1,3}, Xianyuan Liu², Yanhui Sun¹, Shi Sun¹, Hongrui Yan², Wencheng Lu^{2*} and Tianfu Han^{1*}

¹ Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China, ² Heihe Branch, Heilongjiang Academy of Agricultural Sciences, Heihe, China, ³ Council for Scientific and Industrial Research (CSIR)-Oil Palm Research Institute, Kade, Ghana

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Zhihui Shan,
Oil Crops Research Institute
(CAAS), China
Shouping Yang,
Nanjing Agricultural University, China

*Correspondence:

Tianfu Han
hantianfu@caas.cn
Wencheng Lu
hhlwc@sina.com

[†]These authors have contributed
equally to this work

Specialty section:

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

Received: 14 April 2022

Accepted: 01 June 2022

Published: 30 June 2022

Citation:

Jia H, Liang X, Zhang L, Zhang J,
Sapey E, Liu X, Sun Y, Sun S, Yan H,
Lu W and Han T (2022) Improving
Ultra-Low Temperature Preservation
Technologies of Soybean Pollen for
Off-Season and Off-Site Hybridization.
Front. Plant Sci. 13:920522.
doi: 10.3389/fpls.2022.920522

Preserving viable pollen is of great interest to breeders to maintain desirable germplasm for future inbreeding. Ultra-low temperature preservation of pollen is an effective and safe way for long-term storage of plant germplasm resources. In this study, we improved methods for the preservation of soybean pollen at ultra-low temperature. Soybean flowers at the initially-open stage were collected at 6–10 a.m. during the fully-bloom stage of soybean plants and were dehydrated for 10 h and then frozen and stored at -196 or -80°C . *In vitro* culture experiments showed that the viability of preserved pollen remained as high as about 90%. The off-season (local site Heihe) and off-site (Beijing, after long-distance express delivery from Heihe) hybridization verification was conducted, and no significant difference in true hybrid rate was founded between the preserved pollen and the fresh pollen. The ultra-low temperature preservation technology for soybean pollen could break the spatiotemporal limit of soybean hybridization and facilitate the development of engineered soybean breeding.

Keywords: soybean, pollen, ultra-low temperature preservation, germination rate, off-site hybridization

INTRODUCTION

Artificial hybridization between varieties is still the major method for soybean breeding. As the photo-thermal sensitivity of soybean limits the adaptation of soybean varieties to a wide latitude (Liu et al., 2017), it is difficult to achieve intercrossing between varieties with the diverse geographical origin and different flowering times even under the same sowing date and location (Tyagi and Hymowitz, 2003). As a result, the range of available parents is limited and the genetic basis of bred varieties is very narrow (Polito and Luza, 1988). In the conventional breeding programs, soybean hybridization could be completed only when the flowering of selected varieties is at the same time and location. Therefore, a very short time for hybridization results in low hybridization efficiency. The availability of viable pollen is a prerequisite to facilitate breeding in many species to overcome this time-and-space difficulty (Alba et al., 2011; Gowthami et al., 2021).

At present, ultra-low temperature preservation at -80°C , in liquid nitrogen (LN, -196°C), or LN vapor phase ($-150 \sim -180^{\circ}\text{C}$) are all effective for pollen storage (Lu and Chen, 2003; Zhang et al., 2006; Chen et al., 2013; Yang et al., 2014) and for the extended maintenance of pollen viability (Barnabás and Rajki, 1976; Akihama and Omura, 1986; Li, 1987; Ganeshan and Alexander, 1990; Hu and Guo, 1996; Shi et al., 1996; Wang et al., 2003; Cheng et al., 2007; Li et al., 2010; Zhang et al., 2017; Qiu et al., 2018; Oliveira et al., 2021). Cryo-banks of pollen have been established in Japan (Akihama and Omura, 1986), the United States (Connor and Towill, 1993), Canada (Mercier, 2009), China (Zhang et al., 2009, 2014a, 2017), and India (Engelmann, 2004).

To date, the common procedure of pollen ultra-low temperature preservation is as follows: pollen collection, dehydration, freezing, thawing, and field application (Li, 1987; Polito and Luza, 1988; Liu and Wang, 2001). Pollen collection, dehydration methods, and thawing methods directly determine the success or failure of pollen preservation (Liu and Wang, 2001; Shang et al., 2018). To achieve the ideal preservation results, attention should be paid to improving the technology for the preservation of different species and tissues.

The status of pollen for collection is an important factor for pollen preservation (He et al., 2017). The plant pollen viability varies greatly in different developmental stages, which was mainly related to the maturity of pollen (Liang et al., 1993). Marchant et al. (1992) compared the Chinese rose pollen in different open states of flowers and found that pollen of unopened flowers was less likely to be contaminated and more suitable for collection.

For pollen storage, moderate dehydration is also the key to keeping post-preservation pollen viability (Sauve et al., 2000). As temperature decreases, the water inside the cell freezes to form ice crystals, which damage cell membranes and organelles under ultra-low temperature (Towill, 1981; Liu et al., 2015). To reduce or avoid the damage, the water content of the tissues should be reduced appropriately to optimal levels before storage (Nepi et al., 2001; Pacini and Hesse, 2004). And the optimal water content of pollen for preservation at -80°C and -196°C varies among species, even cultivars. Methods for pollen dehydration include desiccant drying (Hu and Guo, 1996), drying at room temperature (Liu and Wang, 2001), drying under incandescent lamps (Jiang and Gao, 1989), freeze-vacuum drying, vacuum drying, etc. Studies have shown that dehydration at room temperature helps to maintain optimal pollen viability (Rajasekharan and Ganeshan, 1994), whereas high temperature during the dehydration process adversely affects the pollen viability (D'antonio and Quiros, 1987). Drying at room temperature and drying under incandescent lamps both achieved excellent drying results (Zhang et al., 2017).

In addition, thawing methods also influenced the pollen viability after preservation (Li, 1987; Polito and Luza, 1988; Shi et al., 1995; Sauve et al., 2000; He et al., 2017). The process of thawing, water absorption, secondary freezing, and osmotic shock of water damage the cell membrane system (Zhang et al., 2006). Thus, it is very important to improve the thawing method. The methods of thawing pollen include thawing at room

temperature, thawing in a warm water bath, and thawing under running water. When the cryopreserved materials were thawed, the temperature of refreezing is -5 to -10°C (Chen, 1989). A warm water bath at $30\text{--}40^{\circ}\text{C}$ is usually used for thawing so that the pollen quickly passes through the dangerous temperature range (Liu and Zhang, 2004).

Several studies have been reported on the storage of soybean pollen. Chen and Ding (1988) found that more than 50% of the pollen dehydrated with calcium chloride was still viable on the 5th day under low-temperature conditions ($2\text{--}3^{\circ}\text{C}$). Gai et al. (1980) found that soybean pollen with low temperature (-3 to -4°C) and dehydration treatment retained viability of 60% for storage of 14–15 days. Perveen and Khan (2009) found that high germination and viability of soybean pollen were maintained under storage in a freezer (-20 and -30°C), and the highest germination percentage was observed under a freeze drier (-60°C). Tyagi and Hymowitz (2003) investigated the germination rate of preserved pollen from seven cultivated and five wild soybean varieties and found that the range of pollen germination rate was 17–77.8% after storage in LN of 7 days without dehydration before storage. In previous studies, the effect of a single factor on the ultra-low temperature preservation of soybean pollen was studied, however, there is a lack of systematic study. This study was undertaken to investigate the effects of pollen collection, dehydration, freezing, thawing, and hybridization verification on the ultra-low temperature preservation of soybean pollen. It is aimed to develop a simple and efficient preservation protocol and realize the breeding application of preserved pollen through long-term storage and long-distance transportation.

MATERIALS AND METHODS

Materials

The soybean pollen used for the investigations of preservation technology was collected from a purple flower variety Heihe43, which is the most widely grown soybean variety, in the Heihe City, Heilongjiang Province, China. In the hybridization verification investigation conducted in Heihe, the white flower variety Jinyuan55 was used as the female parent and the fresh pollen was collected from Heihe43. In the hybridization validation carried out in Beijing, a local white flower variety Zhonghuang39 was used as the female parent. Considering that Heihe43, an elite variety from the north part of northeast China, grew poorly and produced a limited size and number of flowers in Beijing, the fresh pollen was taken from a local purple flower variety Zhonghuang30, an elite variety at Huang-Huai-Hai Valley, to further prove the feasibility of the off-site use of pollen preservation technologies.

Methods

Planting Conditions

In Heihe, Heilongjiang province of China, the field experiments were conducted in 2018–2020 in the Experiment Station of the Heihe Branch of Heilongjiang Academy of Agricultural Sciences ($50^{\circ}15'\text{N}$, $127^{\circ}27'\text{E}$). The sowing dates were May 9, May 11, and May 14, respectively in the relative years. Heihe43 and Jinyuan55

were planted in a 1.5 m row, with 0.6 m space between the rows and a space of 0.07 m between the adjacent plants. In Beijing, Zhonghuang30 and Zhonghuang39 were grown in pots at the campus of the Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (40°130'N, 116°330'E). The seeds of Zhonghuang30 and Zhonghuang39 were planted on 15 June 2020 and 22 June 2020, respectively. Twenty pots for each variety and five uniform plants remained in each pot.

Stage Determination for Collecting Soybean Flowers Bearing Highly Viable Pollen

Based on the degrees of opening of the petals, soybean flowers of Heihe43 were categorized into the following four stages (**Figure 1**): (1) Petal-emerging stage: the petals have just emerged from the calyx and their color can be distinguished, but the height is lower than the calyx (**Figure 1A**); (2) Petal-elongated stage: the petals have extended and exceeded the height of calyx, but the banner flap is still closed (**Figure 1B**); (3) Initially-open stage: the petals grow to the maximum height, and there are obvious cracks at the top of the banner petal (**Figure 1C**); (4) Fully-open stage: wing and keel petals are all fully unfolded (**Figure 1D**). We collected the intact flowers at four stages, respectively, for the detection of pollen viability by *in vitro* culture as described in Section Detection of Pollen Viability.

Dehydration Treatments

The freshly collected soybean flowers were 150 g with three replications. Dehydration treatment was carried out using the oven, incandescent lamp, and natural drying, respectively. The details are as follows:

(1) Dehydrated in the oven: The flowers were placed in a plastic pallet and then were put into the electrothermal blowing drying oven (Lichen101-3BS) for continuous dehydration for 14 h. The temperature was set at 35°C and the humidity in the oven was adjusted to 25% by silica gel desiccants. The ratio of the flower and silica gel was about 1:7. The temperature and humidity in the drying oven were monitored with a temperature and hygrometer, respectively.

(2) Dehydrated using incandescent lamps: The flowers were laid on sulfate paper and placed under a 40 W incandescent lamp for 14 h at room temperature of 25°C and humidity of 55%. The

lamps were 15 cm above the flowers, and the light intensity was about 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$.

(3) Dehydrated at room temperature: The flowers were spread on sulfate paper and placed indoors for 14 h at room temperature of 25°C and humidity of 55%.

During the dehydration process, flowers were randomly taken from each of the above three treatments at drying intervals of 2 h, i.e., 0, 2, 4, 6, 8, 10, 12, and 14 h, respectively, and then the measurement of pollen germination rate was conducted by *in vitro* culture as described as in Section Detection of Pollen Viability. Other dehydrated flowers were preserved and used in subsequent experiments.

Determination of Water Content of Soybean Flowers

Before dehydration, the samples were divided into six parts with three repeats, and about 2 g of each sample was weighed by an analytical balance (ZX224ZH), placed in the Petri dishes, and dehydrated in the drying oven as described in Section Dehydration Treatments. The value of flower weight (FW_i) was recorded at drying intervals of 2 h, i.e., 0, 2, 4, 6, 8, 10, 12, and 14 h drying, respectively. The dry weight (DW) was recorded when it was dehydrated to constant weight in a drying oven at 110°C. The water content (WC_i , %) was computed using the following formula:

$$WC_i = \frac{FW_i - DW}{FW_i} \times 100\%$$

WC_i (%) represents the water content of the flower at i h, FW_i represents the flower weight at i h, and DW represents the final DW.

Freezing and Preservation Methods of Soybean Flowers

Flowers were taken from the drying oven with a temperature of 35°C and humidity of 25% at intervals of 2 h, and each sample was about 1 g in weight, with three replications. The samples were wrapped with the tin foil tightly and then followed by three freezing and preservation treatments: (1) Freezing and preserving in LN; (2) Freezing and preserving in a -80°C freezer (Haier DW-86L338); (3) Freezing in LN and then transferring to a -80°C freezer: put flowers in LN for 72 h and then transfer

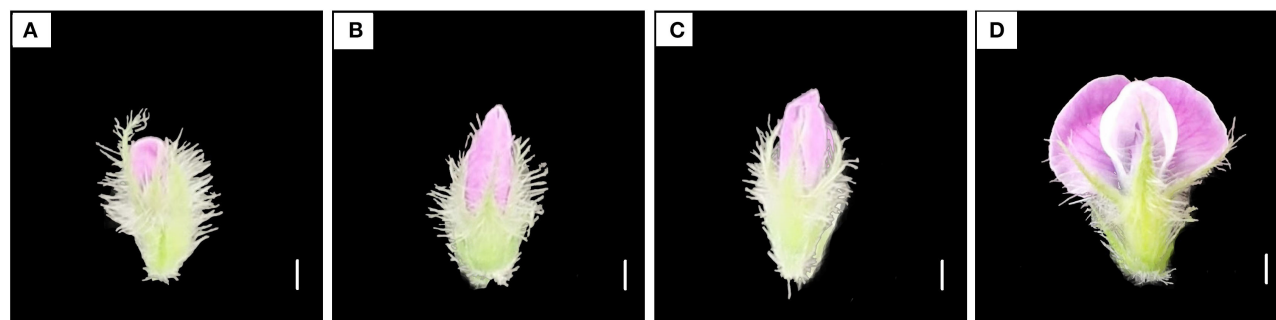


FIGURE 1 | Flower characteristics at four developmental stages. **(A)** Petal-emerging stage; **(B)** Petal-elongated stage; **(C)** Initially-open stage; **(D)** Fully-open stage. Scale bar, 1 mm.

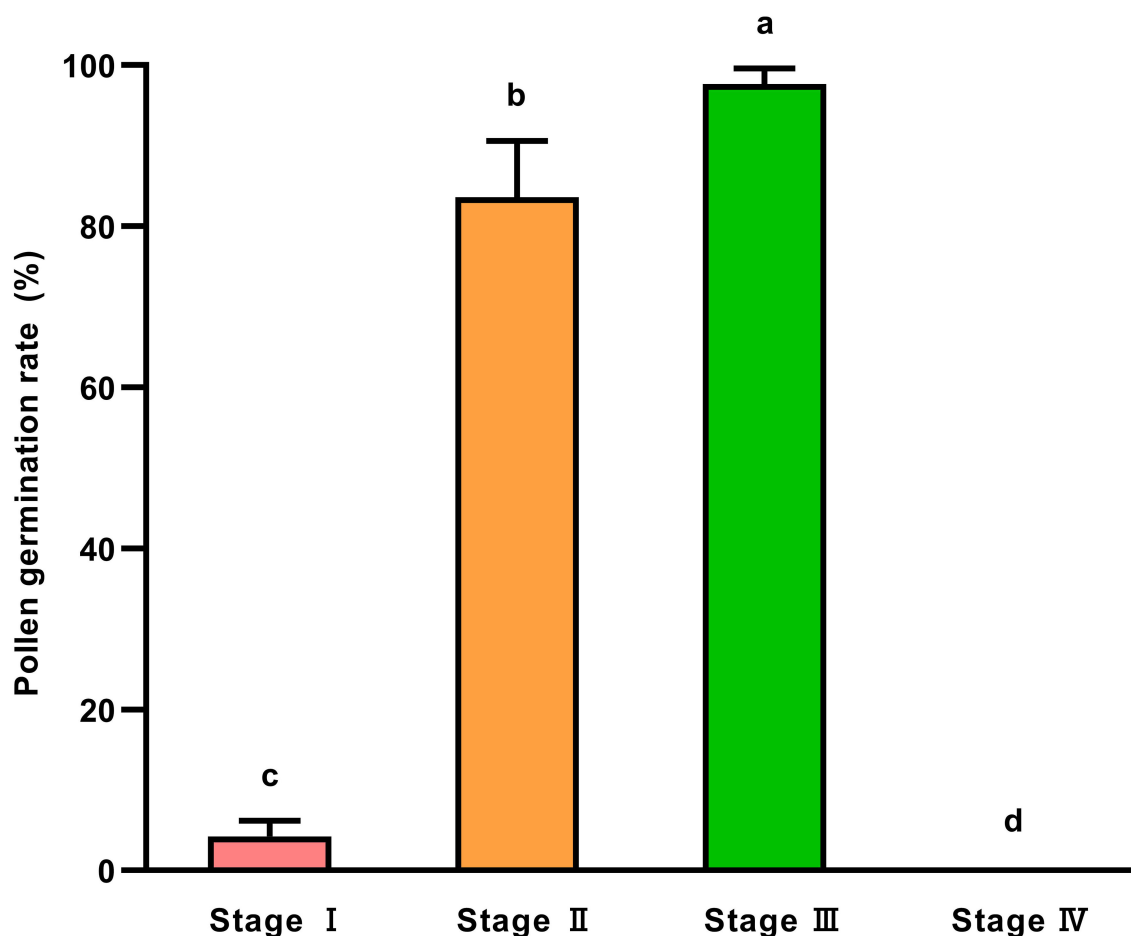


FIGURE 2 | Germination rate of fresh pollen at different developmental stages of soybean flowers. Stage I: Petal-emerging stage; Stage II: Petal-elongated stage; Stage III: Initially-open stage; Stage IV: Fully-open stage. All data were represented as mean \pm SD of three replicates. Welch's and Dunnett's T3 methods were used to compare the means of germination rates of fresh pollen at different developmental stages of soybean flowers ($P < 0.05$).

them to -80°C freezer to preserve. After 1 week, the flower was taken out, thawed, and the pollen germination rate was assessed by *in vitro* culture as described in Section Detection of Pollen Viability.

Thawing Methods of Soybean Flowers

After taking out the tin foil package and wrapping the flowers from the preservation condition described in Section Freezing and Preservation Methods of Soybean Flowers, the samples in the package were thawed for 1 min with the following three methods, respectively: (1) thawing under running water at 20°C ($\pm 5^{\circ}\text{C}$); (2) the package was put in tubes and thawed in the water bath of 35°C ; and (3) thawing at room temperature of 25°C ($\pm 2^{\circ}\text{C}$). The experiment was repeated three times, with six flowers for each repeat selected to measure pollen germination rates by *in vitro* culture as described in Section Detection of Pollen Viability.

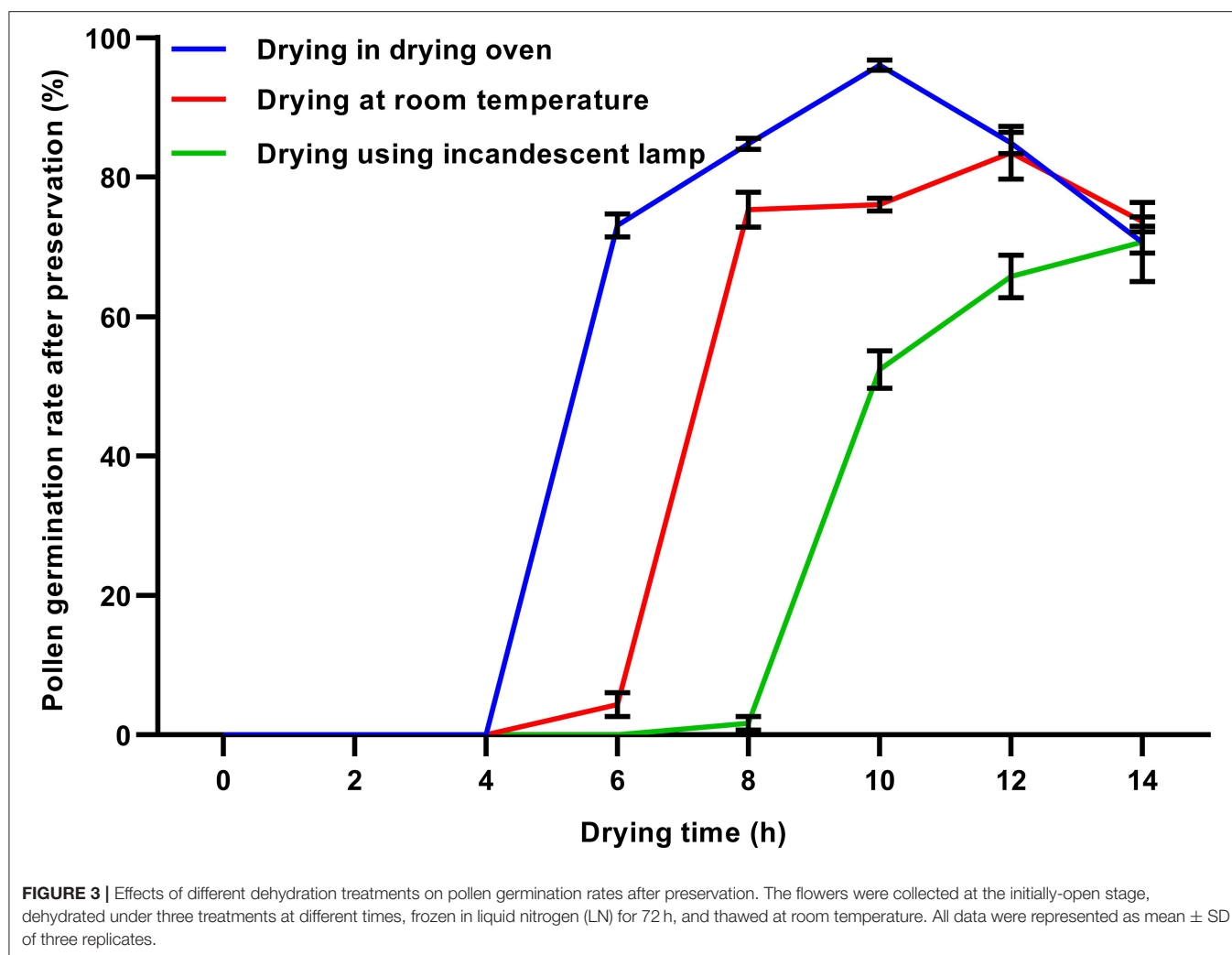
Detection of Pollen Viability

In vitro Culture

The culture medium was prepared using the methods reported by Wang et al. (2016) with some modifications: 19.2% (wt%) sucrose, 68.9 mg/L GA_3 , 0.015% (wt%) H_3BO_3 , 0.05% (wt%) CaCl_2 , 7.5% (wt%) PEG-4000, and the solvent was water. The pollen was cultured in the *in vitro* medium for 20 min at room temperature ($25 \pm 2^{\circ}\text{C}$). The average pollen germination rate was observed under microscopes (Murzider-D106B) randomly and calculated using the data from six fields of six flowers (one for each flower) with three replications. Pollen was scored as germinated if the length of the pollen tube was over two times its diameter, using the formula as indicated below:

$$GR = \frac{b}{a} \times 100\%$$

GR: Germination rate; a : the total number of pollen in each field under microscopes; b : the number of the germinated pollen in each field under microscopes.



Hybridization Verification

Off-season hybridization: In 2018, the flowers of Heihe43 at the initially-open stage were collected in Heihe, dehydrated for 10 h in the drying oven, wrapped with tin foil, frozen in LN for 72 h, and then transferred to a -80°C freezer for 1 year from 15 July 2018 to 24 July 2019. The off-season hybridization was conducted in July 2019 in Heihe. The preserved pollen was taken out in batches from the freezer and thawed at room temperature, and the white flower variety Jinyuan55 (female parent) was pollinated with the preserved pollen. Freshly collected flower at the initially-open stage from Heihe43 in 2019 was used as control and pollination was made synchronously.

Off-site hybridization: On 15 July 2020, flowers of Heihe43 at the initially-open stage were collected in Heihe, dehydrated for 10 h and 14 h in the oven, and were wrapped and frozen in LN for 72 h, and then transferred to a -80°C freezer. After 31 days of preservation, the samples were placed inside a rigid foam plastic box containing 15 kg of dry ice and sent to Beijing (over 1,800 km away from Heihe) by express delivery.

After 3 days, it arrived in Beijing and was stored in the -80°C freezer. From 21 to 25 August 2020, the preserved pollen was taken out in batches from the freezer and thawed at room temperature before use. Pollination was carried out using Zhonghuang39 as the female parent. Fresh pollen of purple flower variety Zhonghuang30 was used as control and pollinated synchronously.

Among the F1 individuals, the plants with purple flowers were regarded as true hybrids and those with white flowers were regarded as false hybrids. The true hybrid rate was calculated by dividing the number of true hybrids by the total number of hybrid plants.

Statistical Analysis

The IBM SPSS 19S Statistics software was used to calculate the descriptive statistics, including the test for homogeneity of variance and one-way ANOVA. Means were separated by Duncan's multiple range test at $P < 0.05$, where the F -test was significant. Welch's and Dunnett's T3 methods were used to compare the means of different treatments for unequal variances.

RESULTS

Germination Rate Differences of Fresh Soybean Pollen at Different Developmental Stages of Flowers

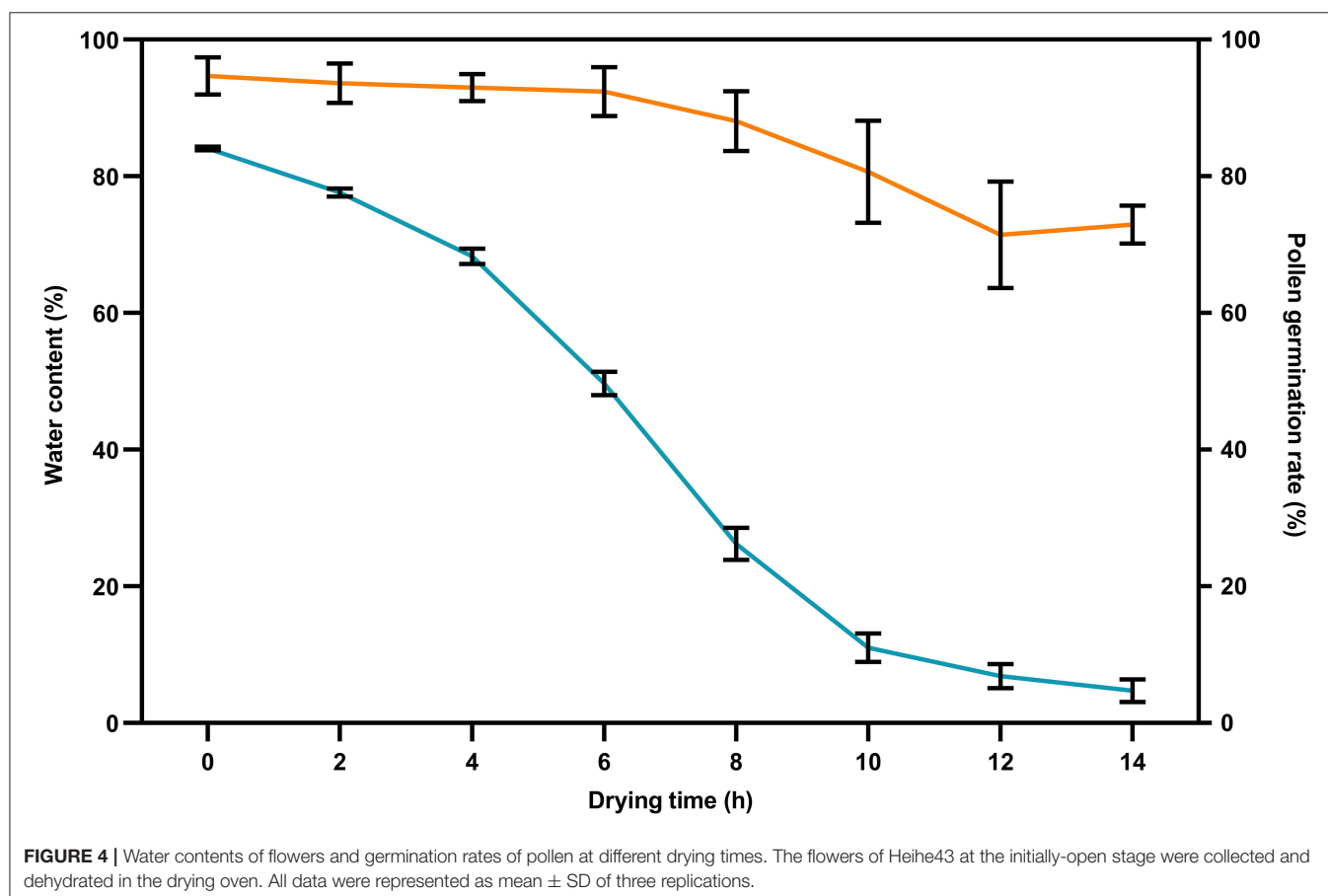
We collected intact flowers at four stages, respectively, for the detection of pollen viability by *in vitro* culture as described in Section Detection of Pollen Viability. The results showed that the average germination rate of fresh pollen collected from intact flowers of Heihe43 was as low as 4.26% at the petal-emerging stage, reaching 83.5 and 97.6% at petal-elongated and initially-open stages, respectively, and failed to germinate during fully-open stage (Figure 2). It suggested that pollen was highly viable during initially-open stages but rapidly declined after the flower fully opened.

Effects of Dehydration Treatments on Soybean Pollen Viability After Ultra-Low Temperature Preservation

To verify the effects of different dehydration methods on pollen viability after ultra-low temperature preservation, the fresh pollen was dehydrated, respectively, in the oven, at room temperature, and under incandescent lamps before preservation. The results showed that within 4 h of dehydration,

the dehydrated pollen in three dehydration treatments failed to germinate after preservation in LN (Figure 3). When the dehydration time was 6 h, the germination rates of soybean pollen dehydrated in an oven, at room temperature, and under an incandescent lamp were 70.1, 4.9, and 0%, respectively, after preservation in LN (Figure 3). With the extension of drying time, the post-preservation pollen germination rates in three dehydration treatments showed an upward trend. The highest post-preservation pollen germination rates in three pre-freezing dehydration treatments were found in the oven drying for 10 h (96.04%), at room temperature for 12 h (83.26%), and under an incandescent lamp for 14 h (70.72%), respectively. Drying in an oven was the fastest way to achieve the high post-preservation pollen germination rate among three dehydration methods.

Low-water content was a precondition for ultra-low temperature preservation of pollen. In the current study, the water content of freshly collected soybean flowers was as high as 84.49% (Figure 4). With the extension of drying time in the oven, the water content of the flower dropped to 41.65% for 6 h, 21.28% for 8 h, 11.0% for 10 h, 6.86% for 12 h, and 4.72% at 14 h (Figure 4). The average germination rate of soybean pollen dehydrated in the oven was above 90% within 6 h, 80.68% at 10 h, and 72.93% at 14 h (Figure 4), indicating that the 10-h dehydration in the oven could reduce the water content of pollen to an optimal range for ultra-low temperature preservation.



Effects of Freezing and Preservation Methods on Germination Rates of Preserved Pollen

To determine the suitable ultra-low temperature preservation methods, the influence of freezing and preservation methods on pollen viability was investigated. The post-preservation germination rate of the pollen dehydrated in the oven was detected under three freezing and preservation treatments, including (1) frozen and preserved in LN; (2) frozen and preserved in -80°C freezer; and (3) frozen in LN for 72 h and then preserved in -80°C freezer. The results showed that within 4 h of dehydration, the average post-preservation germination rate of pollen frozen in LN and then stored in a -80°C freezer was $<20\%$, whereas those of pollen frozen and stored in LN and those frozen and stored in a -80°C freezer were both 0 (Figure 5). When the drying time was 6 h, the post-preservation germination rate of pollen frozen in LN and then preserved in a -80°C freezer was 16.26%, those which were frozen and preserved in LN was 73.04%, and those which were frozen and preserved in a -80°C freezer was 3.26%. When the drying time was 10 h, the average post-preservation pollen germination rate

of all three treatments reached the peak, at about 87.08% (frozen and preserved in a -80°C freezer), 91.08% (frozen in LN for 72 h and then preserved in a -80°C freezer), and 95.95% (frozen and preserved in LN), respectively (Figure 5). The germination rate of preserved pollen with dehydration for 12–14 h showed a decreasing trend but was still higher than 70%, which could meet the demand for soybean hybridization.

Effects of Thawing Methods on Germination Rates of Preserved Pollen

To select the suitable thawing method for soybean pollen after ultra-low temperature preservation, the flower samples wrapped with the tin foil packages were thawed in the following three methods: thawing in a water bath at 35°C , at room temperature, and under running water, respectively. The results showed that the germination rates of preserved pollen thawed under the running water and in a 35°C water bath were significantly higher than those thawed at room temperature ($p < 0.05$) regardless of the freezing and preservation methods (Figure 6). Even that, the pollen germination rates by thawing at room temperature was over 75%, which was high enough for hybridization.

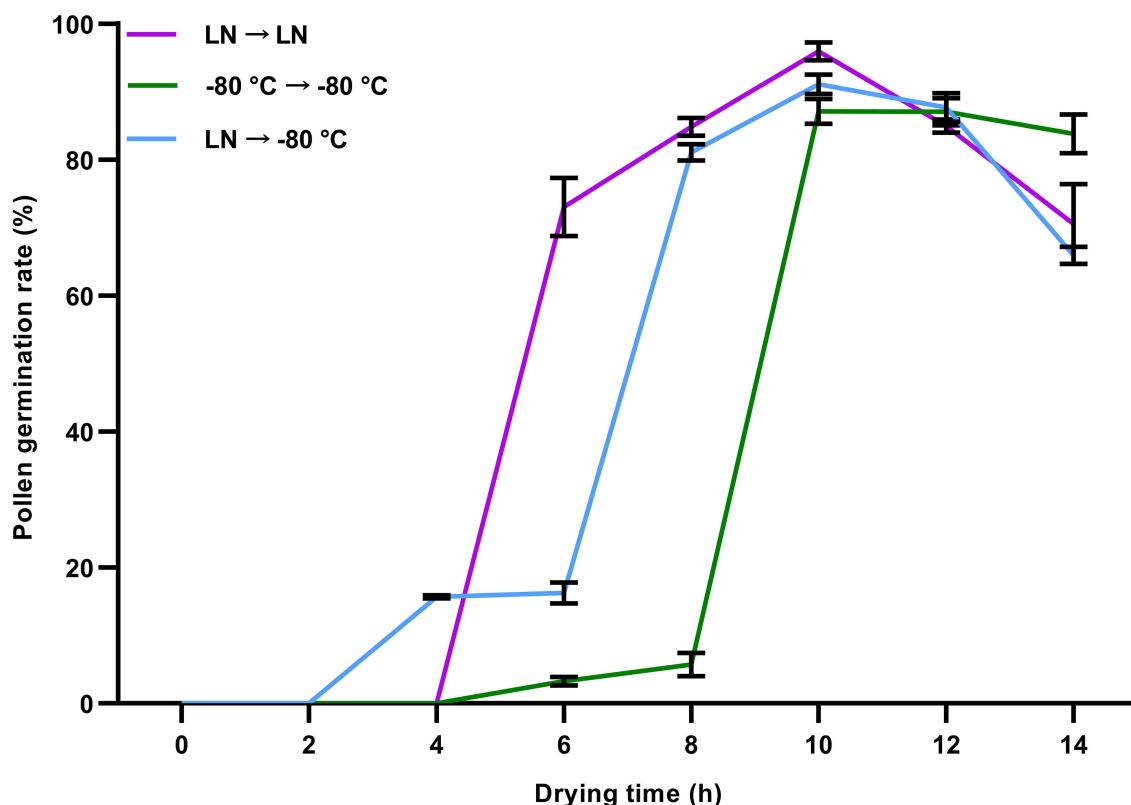


FIGURE 5 | Germination rates of soybean pollen after different freezing and preservation treatments. LN → -80°C , flowers were frozen in liquid nitrogen (LN) for 72 h and then stored in a -80°C freezer; -80°C → -80°C , flowers were frozen and preserved in a -80°C freezer; LN → LN, flowers were frozen and preserved in LN. The flowers at the initially-open stage were collected and dehydrated in a drying oven at different times. After preservation, the flowers were thawed at room temperature for germination tests. All data were represented as mean \pm SD of three replicates.

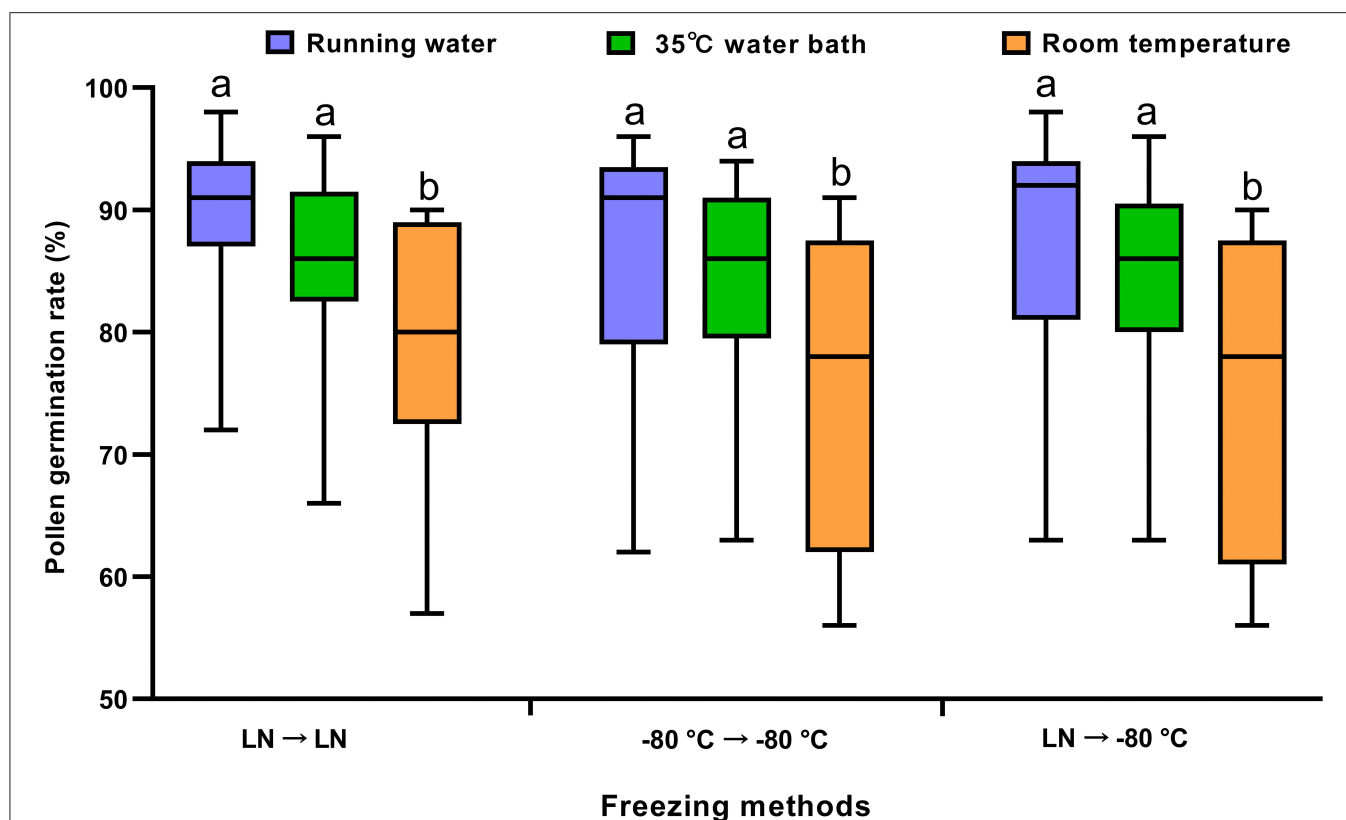


FIGURE 6 | Comparison of germination rates of preserved soybean pollen after different preservation and thawing methods. LN → -80°C , frozen in liquid nitrogen for 72 h and then stored in a -80°C freezer; $-80^{\circ}\text{C} \rightarrow -80^{\circ}\text{C}$, frozen and preserved continuously in a -80°C freezer; LN → LN, frozen and preserved in liquid nitrogen. The preserved samples were thawed for 1 min by three methods. Means of germination rates of preserved pollen after different preservation and thawing methods were separated by Duncan's multiple range test at $P < 0.05$, where the F -test was significant.

Application of Preserved Pollen in Soybean Hybridization

Off-Season Hybridization Application of Preserved Pollen in Heihe

To verify the viability of 1-year preserved pollen of Heihe43 at -80°C , the hybridization experiment was conducted in 2019. Both the preserved and fresh pollen were used to pollinate Jinyuan55 as the female parent. The results showed that there was no significant difference in the true hybrid rate of F1 generation between being pollinated with the preserved pollen for 1 year and the fresh pollen (Figure 7), demonstrating that ultra-low temperature preservation well maintained the viability of the soybean pollen.

Off-Site Hybridization Application of Preserved Pollen in Beijing

To test the feasibility of the use of preserved pollen in other regions, the preserved pollen (flowers) was transported to Beijing, which is 1,800 km away from Heihe, and the hybridization verification tests were carried out in Beijing in July 2020. The results showed that the pollination with preserved pollen resulted

in an even higher true hybridization rate than that using fresh pollen of local elite variety Zhonghuang30 (Table 1).

DISCUSSION

Optimal Developmental Stages of Soybean Flowers for Pollen Collection

Pollen germination rate varies at different developmental stages of flowers in plants. For *Ziziphus* moss, the germination rate of nearly and fully mature pollen was higher than young pollen (Liang et al., 1993). For dogwood (*Cornus florida*), pollen collected from flowers with dehiscence anther was suitable to store (Sauve et al., 2000). In this study, the viability of pollen collected from the flowers at the initially-open stage was higher (Figure 2). The appropriate time for collecting soybean flowers at Heihe was 6–10 a.m. The morphological characteristics of soybean flowers that bore high-viability pollen were determined, ensuring the sampling consistency and high germination rate of pollen. In addition, collecting intact soybean flowers avoided pollen contamination, improved the efficiency of pollen collection, and facilitated the large-amount preservation of pollen at ultra-low temperature.

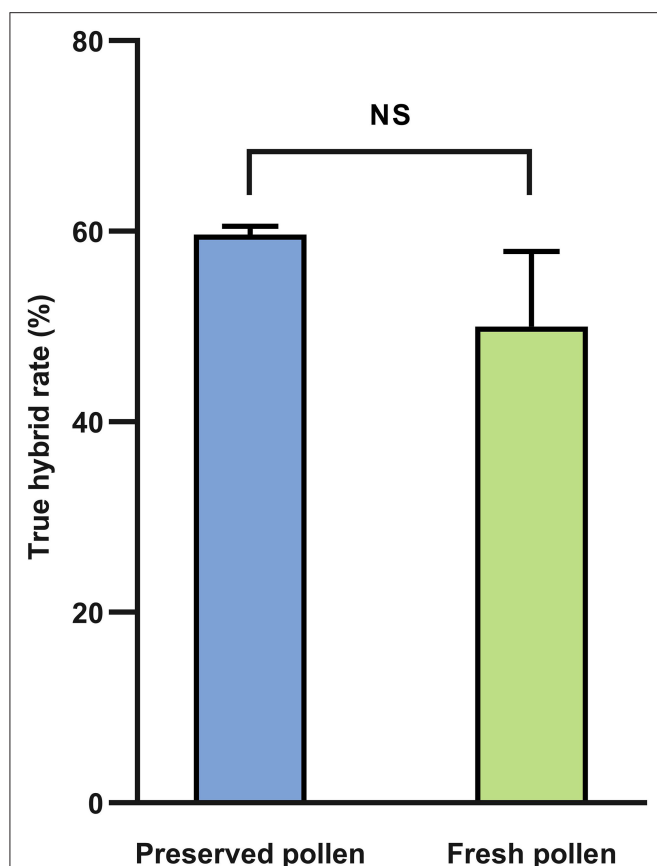


FIGURE 7 | Comparison of the true hybrid rates in F1 generation of combinations using preserved and freshly collected flowers as the pollen sources. Preserved pollen: the flowers of a purple flower variety Heihe43 at the initially-open stage were collected, dehydrated, and frozen in liquid nitrogen (LN) for 72 h and then stored at -80°C , from 15 July 2018 to 24 July 2019; Fresh pollen: Fresh flowers of Heihe 43 at the initially-open stage was collected in Heihe within 0.5 h to pollinate before hybridization. Data were represented as mean \pm SD of two replicates. Means of the true hybrid rates in F1 generation of combinations using preserved and freshly collected flowers as the pollen sources were separated by Duncan's multiple range test at $P < 0.05$, where the F -test was significant.

Selection of Dehydration, Freezing, Preservation, and Thawing Methods for Soybean Pollen

The survival of plant cells after freezing and ultra-low temperature preservation is dependent on the lethal effects of intracellular ice. It is necessary to reduce the water content of pollen to prevent cell damage. Optimal water content is a major factor and varies among species (Zhang et al., 2006; Zhou et al., 2013; Liu et al., 2015; Yu et al., 2019). For litchi pollen, the suitable dehydration method was the use of an air-blowing electric dryer at 35°C for 6 h (Wang et al., 2015). In this study, the water content of freshly collected soybean flowers was around 84% (Figure 4), and the pollen hardly germinated when directly stored in LN without dehydration. It was found that soybean pollen dehydrated in the drying oven presented

germination capacity better than at room temperature and under the incandescent lamp, and 10 h was the optimal drying time for preservation of pollen (Figure 3).

Besides dehydration, pollen viability can be affected by cooling speed, i.e., slow vs. fast cooling, during the preservation procedure (Dinato et al., 2020). A combination of fast-drying and cooling rates may enable the survival of pollen due to the reduction of the time of exposure to dehydration-related deleterious biochemical changes and the inhibition of intracellular ice-crystal formation (Impe et al., 2022). In this study, the 10 h-dehydrated pollen in the drying oven was suitable for all these three freezing and preservation methods, and the 8 h-dehydrated pollen in the drying oven was suitable for preservation in LN, and also suitable for freezing in LN and then preservation in a -80°C freezer (Figure 5). For preservation in LN, it is necessary to maintain the stable LN supply and special container, while the -80°C preservation can be conducted in a -80°C freezer. The preference of freezing and preservation at -196°C , freezing at -196°C , and preservation at -80°C would be up to the preservation conditions available and the ultimate goals for preservation.

Optimal thawing treatments are to avoid ice crystallization in the cells and to prevent the osmotic shock of water during thawing and water absorption from damaging the cell membrane system (Zhang et al., 2006). The suitable temperature range for ice crystallization is $-3 \sim -50^{\circ}\text{C}$ (Liu et al., 2015). If the temperature rises slowly during the thawing treatment and the pollen is in the "dangerous temperature zone" for a long time, the free water will recrystallize, which damages the cells and reduces the pollen vitality. Wang et al. (2003) found that thawing in a gradual treatment of -20°C (12 h) \rightarrow 4°C (12 h) \rightarrow 25°C (12 h) had better potato pollen viability than thawing in a continuous $35\text{--}40^{\circ}\text{C}$ water bath. The thawing effect in running water and warm water bath had no differences and were better than at room temperature (Zhang et al., 2006). This is consistent with our results. Although the viability of pollen thawing at room temperature is slightly lower, it also meets the pollination requirements. Therefore, we selected the thawing method according to the specific situation in breeding.

Off-Season and Off-Site Applications of Preserved Soybean Pollen in Breeding

Pollination is the most direct and effective method to evaluate the viability of preserved pollen (Liu et al., 2015). In this study, the hybridization experiments were carried out locally but off-season in Heihe and off-sited in Beijing, respectively. Hybridization results are greatly affected by many factors, such as temperature, air humidity, the expertise of the person carrying out the hybridization, etc. To ensure the reliability of the hybridization, the crossing in both Heihe and Beijing were carried out in uniform environments and conducted each by a skilled person in soybean hybridization. In Heihe, the true hybrid rate of the F1 generation of being pollinated with the preserved pollen for 1 year was as high as 59.68% (Figure 7), demonstrating that ultra-low temperature preservation maintained the viability of the soybean pollen at least for 1 year.

TABLE 1 | Effects of preserved pollen on off-site pollination in Beijing.

Female parent		Male parent			Pre-treatment of flowers	Number of plants	True hybrid number	True hybrid rate (%)
Variety	Site	Variety	Pollen collection site	Pollen collection date				
Zhonghuang39	Beijing	Heihe43	Heihe	15th July	Dehydrated for 10 h	30	27	84.4
Zhonghuang39	Beijing	Heihe43	Heihe	15th July	Dehydrated for 14 h	15	14	93.3
Zhonghuang39	Beijing	Zhonghuang30	Beijing	21st August to 25th August	No (fresh)	5	3	60.0

On 15 July 2020, pollen of Heihe 43 at the initially-open stage was collected, dehydrated, frozen in LN for 72 h, and then stored at -80°C . On 16 August 2020, the package was sent to Beijing by express delivery. From 21–25 August 2020, hybridization verification tests were carried out in Beijing using the preserved pollen sent from Heihe by express delivery.

The maintenance of pollen viability using ultra-low temperature preservation has been realized in many species. It was reported that the viability of 8–10-years cryopreserved pollen from 12 species/cultivars of ornamental plants was higher than the fresh pollen, 17 species/cultivars retained the same viability as the control, and the viability of pecan pollen was still significantly higher than that of fresh pollen after cryopreservation (preservation in LN) of 13 years (Sparks and Yates, 2002; Ren et al., 2019). In maize, there was no significant difference in pollen pollination capacity between 1- and 2-years cryopreserved pollen and the control (Shi et al., 1996). In pecan and sweet cherry, pollen maintained over 50% viability when stored in a -80°C freezer for 1 year (Ozcan, 2020; Wang et al., 2021). The viability of cryopreserved pollen in some species or cultivars showed a decreasing trend during the cryopreserved process (Ren et al., 2019). The differences in the storage tolerance of pollen in diverse species/cultivars may be also attributed to the pollen type, pollen size, plant taxonomy, etc., besides the pretreatments of preservation. It was known that the binucleate pollen is tolerant to dehydration and has greater viability when compared to trinucleate pollen (Dinato et al., 2020). Soybean pollen is binucleate (Albertsen and Palmer, 1979), and the pollen longevity might be longer by using the improved ultra-low temperature preservation technologies in this study. The monitoring of the preserved pollen and the stress response of preserved pollen under ultra-low temperature was investigated.

In this study, the germination rate of pollen after ultra-low temperature preservation was higher than that of pollen before freezing (Figures 4, 5), and the true hybridization rate of preserved pollen was higher than fresh pollen as well (Figure 7, Table 1). Similar findings were previously reported in other crops (van der Walt and Littlejohn, 1996; Liu et al., 2001). The reason is still not well-understood (Zhang et al., 2006, 2009; Li et al., 2010).

In the breeding programs under the field conditions, parents of a cross must synchronize during flowering time, which limits the scope of parent selection and leads to a narrow genetic base of new varieties. Soybean is a short-day plant, and the flowering time can be adjusted through artificial photoperiod treatments, sowing-date change, and other

measures. However, the photoperiod treatment was costly, and under short-day conditions, soybean plants grew poorly, and produce flowers with a limited size and number of flowers (Zhang et al., 2014b), which was not conducive to pollen production and pollination. Date-of-planting was time-consuming and limited to narrow parents with similar maturity groups (Song et al., 2019). Off-season and off-site applications of preserved pollen allow the collection of flowers at any time and long-term preservation for backup, breaking through the asynchronous flowering barrier between parents belonging to different maturity groups, which will greatly broaden the range of parents for soybean breeding, expand the genetic base of new varieties, and facilitate the engineered breeding of soybean.

CONCLUSION

In this study, we improved ultra-low temperature preservation technologies of soybean pollen for off-season and off-site hybridization. The procedures can be summarized as follows:

- (1) Collect the intact soybean flowers at initially-open stages in the morning (6–10 a.m.).
- (2) Dry the soybean flowers in a drying oven with a humidity of about 25% and a constant temperature of 35°C for 10 h.
- (3) Wrap the dehydrated flower with tin foil, and select either one of the following methods for freezing and preservation: one is to freeze and preserve in LN (LN) for storage, and the other is to freeze in LN and then preserve in freezers at -80°C .
- (4) Put the preserved flowers into a container filled with enough dry ice, and transport them to the hybridization site by express delivery.
- (5) Thaw the foil-wrapped flowers at room temperature for pollination application.

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

XL, HJ, and LZ conducted the major experimental works. XL and YS joined the experiments. JZ, SS, and HY provided advice and technical support. HJ, XY, TH, JZ, LZ, and ES wrote the manuscript. TH and WL coordinated and supervised the project. All authors read and approved the final manuscript.

REFERENCES

- Akihama, T., and Omura, M. (1986). "Preservation of fruit tree pollen," in *Trees I. Biotechnology in Agriculture and Forestry*, ed Y. P. S. Bajaj (Berlin: Springer), 101–112. doi: 10.1007/978-3-642-70576-2_7
- Alba, V., Bisignano, V., Alba, E., Stradis, A., and Polignano, G. (2011). Effects of cryopreservation on germinability of olive (*Olea europaea* L.) pollen. *Genet. Resour. Crop. Evol.* 58, 977–982. doi: 10.1007/s10722-011-9736-z
- Albertsen, M., and Palmer, R. (1979). A comparative light and electron-microscopic study of microsporogenesis in male sterile (ms1) and male fertile soybeans [*Glycine max* (L.) Merr.]. *Am. J. Bot.* 66, 253–263. doi: 10.1002/j.1537-2197.1979.tb06222.x
- Barnabás, B., and Rajki, E. (1976). Storage of maize (*Zea mays* L.) pollen at -196°C in liquid nitrogen. *Euphytica* 25, 747–752. doi: 10.1007/BF00041614
- Chen, P. (1989). Cryopreservation of plant tissue cultures. *Wuhan Bot. Res.* 7, 390–398.
- Chen, W., and Ding, F. (1988). Hybridization effect of different preservation methods on summer soybean pollen. *Shandong Agr. Sci.* 4, 26–27.
- Chen, X., Zhang, J., Xin, X., Huang, B., and Lu, X. (2013). Progress on cryopreservation state and research of plant germplasm resources. *J. Plant Genet. Resour.* 14, 414–427. CNKI:11-4996/S.20130402.1733.004
- Cheng, W., Zhao, Z., Guo, J., Liu, B., Zeng, J., and Lai, J. (2007). Preliminary report on storage under low temperature of *Betula alnoides* pollen. *J. Zhejiang Forest. Sci. Tech.* 6, 49–52. doi: 10.3969/j.issn.1001-3776.2007.06.012
- Connor, K., and Towill, L. (1993). Pollen-handling protocol and hydration/dehydration characteristics of pollen for application to long-term storage. *Euphytica* 68, 77–84. doi: 10.1007/BF00024157
- D'antonio, V., and Quiros, C. (1987). Viability of celery pollen after collection and storage. *Hort. Sci.* 22, 479–481.
- Dinato, N., Santos, I., Vigna, B., Paula, A., and Favero, A. (2020). Pollen cryopreservation for plant breeding and genetic resources conservation. *Cryoletters* 41, 115–127.
- Engelmann, F. (2004). Plant cryopreservation: progress and prospects. *In Vitro Cell Dev. Biol. Plant.* 40, 427–433. doi: 10.1079/IVP.2004541
- Gai, J., Hu, Y., Chen, J., and Gu, Y. (1980). Experiment on the retaining of pollen viability of soybeans. *Acta Agron. Sin.* 6, 11–16.
- Ganeshan, S., and Alexander, M. (1990). Fertilizing ability of cryopreserved grape (*Vitis vinifera* L.) pollen. *Vitis* 29, 145–150.
- Gowthami, R., Sharma, N., Gangopadhyay, K., Rajkumar, S., Pathania, P., and Agrawal, A. (2021). Cryopreservation of pollen of *Abelmoschus Moschatus* Medik. subsp. *Moschatus* as an aid to overcome asynchronous flowering for wide hybridization with cultivated okra [*A. esculentus* (L.) Moench]. *Cryoletters* 42, 233–244.
- He, W., Xiao, Q., Pu, G., Huang, X., Li, Y., and Shi, L. (2017). Effect of walnut pollen on 'Shuangzao' fruit quality and early fruit of several. *J. Hunan Agri. Univ.* 43, 266–269. doi: 10.13331/j.cnki.jhau.2017.03.008
- Hu, J., and Guo, C. (1996). Studies on the cryopreservation (-196°C) of pollen of restoring line in hybrid rice. *Acta Agron. Sin.* 22, 72–77.
- Impe, D., Ballesteros, D., and Nagel, M. (2022). Impact of drying and cooling rate on the survival of the desiccation-sensitive wheat pollen. *Plant Cell Rep.* 41, 447–461. doi: 10.1007/s00299-021-02819-w
- Jiang, Y., and Gao, Z. (1989). Ultra-low temperature (-196°C) storage of peach and pear pollens. *Acta Agric. Shanghai* 5, 1–8.
- Li, B., Wang, H., and Liu, Y. (2010). Pollen cryopreservation of Japanese tree peony cultivars. *J. Beijing Forest. Univ.* 32, 297–300. CNKI:SUN:BJLY.0.2010-04-056
- Li, X. (1987). Studies on rice pollen storage. *Bull. Agri. Sci. Tech.* 6, 2–4.

FUNDING

This work was supported by the Key-Area Research and Development Program of Guangdong Province (2020B020220008), the China Agriculture Research System (CARS-04), and the Chinese Academy of Agricultural Sciences Innovation Project.

- Liang, L., Xu, B., Zheng, C., and Zhou, C. (1993). Pollen cryopreservation and pollen protoplast isolation in *Brassica campestris* var. *purpurea*. *J. Integr. Plant Bio.* 35, 733–738.
- Liu, L., Cao, C., and Liu, X. (2015). Research progress in cryopreservation of cucurbit pollen. *China Cucu. Veget.* 28, 1–4. doi: 10.3969/j.issn.1673-2871.2015.02.001
- Liu, X., Wu, J., Ren, H., Qi, Y., Li, C., and Cao, J. (2017). Genetic variation of world soybean maturity date and geographic distribution of maturity groups. *Breed. Sci.* 67, 221–232. doi: 10.1270/jsbbs.16167
- Liu, Y., and Wang, X. (2001). Advances in the study of cryopreservation of fruit tree germplasm. *Life Sci. Res.* 5, 227–232. doi: 10.16605/j.cnki.1007-7847.2001.s1.047
- Liu, Y., and Zhang, Y. (2004). Pollen cryopreservation of *Prunus mume*. *J. Beijing Forest. Univ.* 5, 22–25. doi: CNKI:SUN:BJLY.0.2004-S1-005
- Liu, Y., Zhou, H., and Fang, B. (2001). Cryopreservation of seed of ornamental plants. *J. Beijing Forest. Univ.* 23, 39–44. doi: 10.13332/j.1000-1522.2001.04009
- Lu, X., and Chen, X. (2003). Progress of conservation and research of crop germplasm resources in China. *Sci. Agric. Sin.* 36, 1125–1132. doi: CNKI:SUN:ZGNX.0.2003-08-000
- Marchant, R., Power, J., Davey, M., Chartier-Hollis, J., and Lynch, P. (1992). Cryopreservation of pollen from two rose cultivars. *Euphytica* 66, 235–241. doi: 10.1007/BF00025309
- Mercier, S. (2009). The role of a pollen bank in the tree genetic improvement program in Québec (Canada). *Grana* 34, 367–370. doi: 10.1080/00173139509429468
- Nepi, M., Franchi, G., and Padni, E. (2001). Pollen hydration status at dispersal: cytophysiological features and strategies. *Protoplasma* 216, 171–180. doi: 10.1007/BF02673869
- Oliveira, A., Ledo, A., Polek, M., Krueger, R., Shepherd, A., and Volk, G. (2021). Optimization of *in vitro* germination and cryopreservation conditions for preserving date palm pollen in the USDA National Plant Germplasm System. *Plant Cell Tiss. Org. Cult.* 144, 223–232. doi: 10.1007/s11240-020-01907-1
- Ozcan, A. (2020). Effect of low-temperature storage on sweet cherry (*Prunus avium* L.) pollen quality. *HortScience* 55, 258–260. doi: 10.21273/HORTSCI14660-19
- Pacini, E., and Hesse, M. (2004). Cytophysiology of pollen presentation and dispersal. *Flora* 199, 273–285. doi: 10.1078/0367-2530-00156
- Perveen, A., and Khan, S. (2009). Maintenance of pollen germination capacity of *Glycine max* (L.) Merr. (Papilionaceae). *Pak. J. Bot.* 41, 2083–2086.
- Polito, V., and Luza, J. (1988). Low temperature storage of pistachio pollen. *Euphytica* 39, 265–269. doi: 10.1007/BF00037105
- Qiu, Z., Guo, K., Wen, Z., Deng, B., Li, Z., and Wen, X. (2018). Pollen yield and low-temperature preservation of Guizhou main cherry varieties. *Southwest Chin. J. Agri. Sci.* 31, 2185–2190. doi: CNKI:SUN: XNYX.0.2018-10-032
- Rajasekharan, P., and Ganeshan, S. (1994). Freeze preservation of rose pollen in liquid nitrogen: feasibility, viability and fertility status after long-term storage. *J. Hort. Sci.* 69, 565–569. doi: 10.1080/14620316.1994.11516488
- Ren, R., Li, Z., Li, B., Xu, J., Jiang, X., Liu, Y., et al. (2019). Changes of pollen viability of ornamental plants after long-term preservation in a cryopreservation pollen bank. *Cryobiology* 89, 14–20. doi: 10.1016/j.cryobiol.2019.07.001
- Sauve, R., Craddock, J., Reed, S., and Schlarbaum, S. (2000). Storage of flowering dogwood (*Cornus florida* L.) pollen. *Hort. Sci.* 35, 108–109. doi: 10.21273/HORTSCI.35.1.108

- Shang, C., Gao, Z., Hu, X., Wang, Y., Yao, H., Li, X., et al. (2018). Effects of different storage methods on the pollination effect of tomato. *J. Northern Agric.* 46, 113–116. doi: CNKI:SUN:SHZK.0.2016-06-003
- Shi, S., Zhang, Z., and Xiao, J. (1995). Chromosome observation of maize pollen progeny after long-term cryopreservation. *Crop Var. Resour.* 2, 45–47.
- Shi, S., Zhang, Z., and Xiao, J. (1996). Genetic stability in maize pollen after long-term cryopreservation. *Acta Agron. Sin.* 22, 409–413.
- Song, W., Sun, S., Ibrahim, S. E., Xu, Z., Wu, H., and Hu, X. (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
- Sparks, D., and Yates, I. (2002). Pecan pollen stored over a decade retains viability. *Hort. Sci.* 37, 176–177. doi: 10.21273/HORTSCI.37.1.176
- Towill, L. (1981). Liquid nitrogen preservation of pollen from tuber bearing *Solanum* species. *Hort. Sci.* 16, 177–179.
- Tyagi, R., and Hymowitz, T. (2003). Pollen from *Glycine* species survive cryogenic exposure. *CryoLetters* 24, 119–124. doi: 10.1163/156854003765911757
- van der Walt, I., and Littlejohn, G. (1996). Storage and viability testing of protea pollen. *J. Am. Soc. Hortic. Sci.* 121, 804–809. doi: 10.21273/JASHS.121.5.804
- Wang, C., Cao, Y., Zhang, L., Wang, W., Yan, C., Sun, X., et al. (2016). Optimization of main influencing factors of soybean pollen germination *in vitro* by response surface methodology. *Jiangsu Agric. Sci.* 44, 94–98. doi: 10.15889/j.issn.1002-1302.2016.03.025
- Wang, L., Wu, J., Chen, J., Fu, D., Zhang, C., and Cai, C. (2015). A simple pollen collection, dehydration, and long-term storage method for litchi (*Litchi chinensis* Sonn.). *Sci Hortic.* 188, 78–83. doi: 10.1016/j.scienta.2015.03.021
- Wang, X., Wu, Y., and Lombardini, L. (2021). *In vitro* viability and germination of *Carya illinoensis* pollen under different storage conditions. *Sci. Hortic.* 275, 109662. doi: 10.1016/j.scienta.2020.109662
- Wang, Y., Zhang, F., and Wang, D. (2003). The cryopreservation of potato pollen. *Acta Hortic. Sin.* 30, 683–686. doi: 10.3321/j.issn:0513-353X.2003.06.010
- Yang, K., Feng, Z., Zhang, P., and Cheng, T. (2014). Distant hybridization between *Acer* species. *J. Shandong Forest. Sci. Tech.* 44, 21–24. doi: 10.3969/j.issn.1002-2724.2014.02.006
- Yu, Q., Yang, Y., Liu, H., Zheng, X., Kong, F., Zhang, H., et al. (2019). Cryopreservation of *Sorbus pohuashanensis* pollens. *Fujian J. Agri. Sci.* 34, 409–415. doi: 10.19303/j.issn.1008-0384.2019.04.005
- Zhang, J., Lu, X., Xin, X., Yin, G., He, J., and Huang, B. (2017). Cryopreservation of *Citrus* anthers in the national crop genebank of China. *In Vitro Cell Dev. Biol. Plant.* 53, 318–327. doi: 10.1007/s11627-017-9848-z
- Zhang, J., Xin, X., Yin, G., Lu, X., and Chen, X. (2014a). *In vitro* conservation and cryopreservation in national genebank of China. *Acta Hortic.* 309–317. doi: 10.17660/ActaHortic.2014.1039.39
- Zhang, Y., Chen, R., Huang, C., and Liu, Y. (2009). Cryo-banking of *Prunus mume* pollen and its application in cross-breeding. *Cryoletters* 30, 165–170.
- Zhang, Y., Shang, X., and Liu, Y. (2006). Advances in research of pollen cryopreservation. *J. Beijing Forest. Univ.* 28, 139–147. doi: 10.13332/j.1000-1522.2006.04.026
- Zhang, Y., Sun, S., Yang, X., Sun, X., Wu, C., and Han, T. (2014b). Improvement of soybean hybridization success rate during winter nursing in Hainan Island. *Acta Agro. Sin.* 40, 1296–1303. doi: 10.3724/SPJ.1006.2014.01296
- Zhou, Y., Peng, Z., Jiang, X., Wang, K., Wei, X., and Xiao, S. (2013). Study on vitality and cryopreservation of 20 *Spathiphyllum floribundum* pollen. *Chin. Agri. Sci. Bull.* 29, 113–117. doi: 10.3969/j.issn.1000-6850.2013.01.023

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 Jia, Liang, Zhang, Zhang, Sapey, Liu, Sun, Sun, Yan, Lu and Han. 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 Neural Network for Discrimination and Classification of Tropical Soybean Genotypes of Different Relative Maturity Groups

Lígia de Oliveira Amaral^{1*}, Glauco Vieira Miranda², Bruno Henrique Pedroso Val¹, Alice Pereira Silva¹, Alyce Carla Rodrigues Moitinho¹ and Sandra Helena Unêda-Trevisoli¹

¹Laboratory of Biotechnology and Plant Breeding, Department of Agricultural Sciences, São Paulo State University - UNESP/FCAV, Jaboticabal, Brazil, ²Department of Agronomy Coordination, Federal Technological University of Paraná, Curitiba, Brazil

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University,
China

Reviewed by:

Milind Ratnaparkhe,
ICAR Indian Institute of Soybean
Research, India
Javaid Akhter Bhat,
Nanjing Agricultural University, China

*Correspondence:

Lígia de Oliveira Amaral
ligiaoamaral@hotmail.com

Specialty section:

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

Received: 12 November 2021

Accepted: 20 May 2022

Published: 12 July 2022

Citation:

Amaral LO, Miranda GV, Val BHP,
Silva AP, Moitinho ACR and
Unêda-Trevisoli SH (2022) Artificial
Neural Network for Discrimination
and Classification of Tropical
Soybean Genotypes of Different
Relative Maturity Groups.
Front. Plant Sci. 13:814046.
doi: 10.3389/fpls.2022.814046

Soybean has a recognized narrow genetic base that often makes it difficult to visualize available genetic and phenotypic variability and identify superior genotypes during the selection process. However, the phenotypic expression of soybean plants is highly affected by photoperiod and the cultivation of a given variety is performed in the latitude range that presents ideal conditions for its development based on its relative maturity group (RMG) for the optimization of the phenotypic expression of its genotype. Based on the above, this study aimed to evaluate the efficiency of artificial neural networks (ANNs) as a tool for the correct discrimination and classification of tropical soybean genotypes according to their relative maturity group during the population selection process with the aim of optimizing the phenotypic performance of these selected genotypes. For this purpose, three biparental populations were synthesized, one with a wide genetic variability for the RMG character obtained from the hybridization between genitors of maturity groups RMG 5 (Sub-tropical 23° LS) × RMG 9.4 (Tropical 0° LS) and two populations with a narrow variability obtained between genitors RMG 7.3 (Tropical 20° LS) × RMG 9.4 and RMG 5.3 × RMG 6.7, respectively. Criteria for comparing the developed ANN architecture with Fisher's linear and Anderson's quadratic parametric discriminant methodologies were applied to the data for the discrimination and classification of the genotypes. ANN showed an apparent error rate of less than 8.16% as well as a low influence of environmental factors, correctly classifying the genotypes in the populations even in cases of reduced genetic variability such as in the RMG 5 × RMG 6 population. In contrast, the discriminant functions were inefficient in correctly classifying the genotypes in the populations with genealogical similarity (RMG 5 × RMG 6) and wide genetic variability, with an error rate of more than 50%. Based on the results of this study, ANN can be used for the discrimination of genotypes in the initial generations of selection in breeding programs for the development of high performance cultivars for wide and reduced photoperiod amplitudes, even with fewer selection environments, more efficiently, and with fewer time and resources applied. As a result of similarity between the parents, ANN can correctly classify genotypes from populations with a narrow genetic base, in addition to pure lines and genotypes with a high degree of inbreeding.

Keywords: machine learning, photoperiod, *glycine max*, relative maturity, data mining, apparent error rate

INTRODUCTION

Soybean is the most important oilseed in the world and whose genetic improvement plays a significant role in the continuous growth of the crop. Soybean is highly sensitive to the photoperiod and considered a short-day species. The day length at the location of cultivation directly affects plant growth. In addition, the photoperiod influences the change from vegetative to reproductive stage, and consequently, the flowering, total cycle, and grain yield (Garner and Allard, 1930), consequently influencing the entire phenotypic performance in the field. Owing to this influence, the optimum photoperiod conditions for cultivars are restricted to a certain latitude range according to the environment in which the genotypes are cultivated (Lin et al., 2021). In Brazil, cultivars are distributed among 13 relative maturity groups (RMGs) classified geographically based on plant growth and development. Owing to the large territorial extent and latitude variation, Brazil comprises RMGs 5–9, respectively, from the south (latitude 30°) to the north of the country (latitude 0°) (Alliprandini et al., 2009).

The cultivation of a variety in an RMG different from its suitable one may result in undesired cycle elongation or reduction, insufficient or exaggerated vegetative development, susceptibility to pests and diseases specific to a certain time of the year, and low productivity (Miladinović and Đorđević, 2011), impairing their phenotypic performance. Genetically, the time to reach flowering and maturity is controlled by genes *E*. Thus far, 11 major *loci* (*E1–E11* and *J*) involved in the control of these characteristics have been identified in soybean (Samanfar et al., 2017). In general, except for *E6*, *E9*, *E11*, and *J* genes, the dominant allele of *E* genes leads to late flowering and maturity, whereas an increase in the number of recessive alleles entails precociousness of the variety (Lin et al., 2021).

The yield potential of cultivars at their appropriate production locations is maximized. Correct estimation of the phenological stages of the soybean plant allows improvement in the flexibility to modify its development as a whole and use the characteristics controlled by other genes that are affected by the durations of the vegetative and reproductive stages. The number of nodes and pods, growth habit, and characteristics related to the occurrence of higher temperatures at a certain stage of the plant such as oil and protein content and nitrogen and phosphorus concentrations in grains, can be explored well, depending on the objectives of the breeding program (Miladinović et al., 2018).

The effect of a phenotype is the sum of the genetic effects, environmental effects, and their interaction. The evolution of technology and the improvement of methodologies in breeding programs aim to isolate the environmental effects to the maximum extent to increase the efficiency of the selection of genotypes based on the genetic effects. Artificial neural networks (ANNs) have a high capacity in predicting, recognizing, discriminating, and classifying patterns. Moreover, different from parametric statistical approaches, they capture the complex characteristics of a dataset in addition to being slightly susceptible to noise and outliers and being suitable for nonlinearly separable problems common to agricultural experimentation (Kavzoglu and Mather, 2003; Sudheer et al., 2003; Haykin, 2008). Currently,

at the experimental level, ANN models have been used in the prediction of genetic values (Soares et al., 2015), adaptability and stability (do Carmo Oda et al., 2019), phenotyping (Sá, 2018), yield estimates (Lu et al., 2022), genetic diversity (Rahimi et al., 2019; Taratuhin et al., 2020), disease detection, and classification (Hang et al., 2019; Trivedi et al., 2021). Moreover, they have demonstrated that the efficiency in the breeding stages can be increased, which can reduce the time and cost of obtaining high-performance cultivars.

Based on the above, the objective of the present study was to evaluate the efficiency of ANNs as a tool for the correct discrimination and classification of tropical soybean genotypes according to their relative maturity group during the population selection process, with the aim of optimization of the phenotypic performance of these selected genotypes. The possible application of this analysis is expected to obtain high performance cultivars for a wide range of photoperiods in soybean growing regions in Brazil.

MATERIALS AND METHODS

Plant Material

Three soybean populations with different ranges of genetic variability were synthesized and evaluated for their RMG characteristics. Hybridizations were performed at the Department of Agricultural Production Sciences, Faculty of Agricultural Sciences, São Paulo University - UNESP/FCAV located at latitude 21°14'58"S and longitude 48°17'08"W, in Jaboticabal, São Paulo, Brazil. The population with a wide genetic variability for the RMG character was obtained from the hybridization between the genitors BMX Veloz (RMG 5.0; sub-tropical 23° LS) × BRS 278 RR (RMG 9.4; tropical 0° LS), called the Brazil population. The Northern and Southern populations, characterized by the most restricted variability for the character group of relative maturity, were established from crosses between the cultivars BRS 245 RR (RMG 7.3; Tropical 20° LS) and BRS 278 RR (RMG 9.4) and between cultivars BMX Energia (RMG 5.3) and BMX Potência (RMG 6.7), respectively. The controls of each population were their respective genitors, in addition to the cultivars TMG 7262 RR (RMG 6.2), TMG 1174 RR (RMG 7.4), and TMG 1179 RR (RMG 7.9).

The parent cultivar BRS 278 RR was approximately 73 cm tall, had an average cycle of 115–127 days, determined growth habit, brown pubescence, purple flower color, RMG of 9.4, and resistance to lodging, *Xanthomonas axonopodis*, *Cercospora sojina*, and *Diaporthe phaseolorum*, were susceptible to common soybean mosaic, cyst nematode, *Meloidogyne incognita*, and *Meloidogyne javanica*, and tolerant to stem necrosis virus (CpMMV). BMX Veloz was of medium size, had an average cycle of 120 days, indeterminate growth habit, light brown pubescence color, purple flower color, RMG of 5.0, was resistant to *D. phaseolorum* and *Phytophthora sojae* (*RPS1k* gene), and was moderately resistant to *C. sojina* and *X. axonopodis*. BMX Energia was of medium size, had an indeterminate growth habit, gray pubescence color, purple flower color, RMG of 5.3, resistance to *D. phaseolorum*, and moderately resistant to

C. sojina and *X. axonopodis*. BMX Potência had a height of 90 to 100 cm, indeterminate growth habit, gray pubescence color, white flower color, RMG of 6.7, was susceptible to the nematode *M. incognita* and moderately resistant to the nematode *Meloidogyne javanica*, and resistant to lodging, *D. phaseolourum*, *C. sojina*, and *Phytophthora sojae*. BRS 245 RR had a height of 70 to 94 cm depending on the planting region, determined growth habit, brown pubescence color, white flower color, RMG of 7.3, was resistant to *D. phaseolourum*, *C. sojina*, and common soybean mosaic, tolerant to stem necrosis virus, moderately susceptible to powdery mildew, and susceptible to cyst nematodes, *M. incognita*, and *M. javanica*.

All genitor cultivars were transgenic, with resistance to the herbicide glyphosate.

To perform the analyses, the controls were considered as distinct populations and there were 11 populations in total (Table 1).

Experimentation

Jaboticabal is located at latitude 21°15'19"S, presenting ideal photoperiod conditions for RMG genotypes 6–8 originating from the long rainy period in the region from November (spring) to April (autumn), allowing the cultivation of soybean cultivars with a cycle of up to 150 days. Field evaluation data were obtained from years 2017/2018, 2018/2019, and 2019/2020, corresponding to the filial generations from F₃ to F₆ for the Brazil population and F₄ to F₇ for the Northern and Southern populations. The three populations were conducted in the three agricultural years with a variable number of progenies, in addition to four commercial cultivars as controls within each population. The Brazil population consisted of 220 progenies in 2017/2018, 252 in 2018/2019 and 252 in 2019/2020. The Southern population consisted of 120 progenies in 2017/2018, 168 in 2018/2019 and 168 in 2019/2020. The Northern population consisted of 60 progenies in 2017/2018, 60 in 2018/2019 and 104 in 2019/2020. The agronomic characteristics evaluated were: number of days to flowering (NDF), number of days to maturity (NDM), total crop cycle (CYCLE), first pod insertion height (AIV), plant height at maturity (APM), lodging (Ac), agronomic value (VA), and grain yield (PG). Ac was evaluated based on a visual rating scale ranging from 1 (all plants erect) to 5 (all

plants lodged), and VA was evaluated with a visual grading scale ranging from 1 (plants with poor agronomic characteristics) to 5 (plants with excellent agronomic characteristics). All experiments (three populations in three crop years) were conducted in Federer's augmented block design (Federer, 1956), with randomized controls in all experimental blocks. Each experimental plot of each evaluated genotype consisted of a 5-m-long row, with a 0.5-m spacing between rows, and a sowing density of 15 seeds per linear meter. The agronomic characteristics were evaluated on five individual plants within each experimental plot, in all populations and in all agricultural years, making up a robust set of evaluated data.

Discriminant Analyses

Fisher and Anderson discriminant analyses are linear combinations of the observed characteristics that present the best discrimination power among all possible linear combinations of the same characteristics (Johnson and Wichern, 2002). In these methodologies, the total dataset is divided into a training set (80% of the data) responsible for obtaining the discriminant functions and a validation or test set (20% of the data) responsible for validating the functions. For the test set to be a representative sample of the training set, several data partitions were performed, where the mean and variance of each generated set pair were compared, and the pair in which the estimates were the closest possible was selected. Cross-validation of the data and Fisher's linear and Anderson's quadratic discriminant functions were performed in Genes Computer Application (Cruz, 2008) according to the methodology of Cruz et al. (2014).

Analysis by ANN

After several experiments on the best architecture for the multilayer perceptron network type, neural network architecture 12–64–128–11 built in Python 3.6 using Keras as the frontend, TensorFlow 2.3.0 as the backend, and Scikit-learn 0.22.2 was adopted. It was necessary to convert the categorical variable (year) to one-hot representation. Thus, the number of input neurons was 12, corresponding to the POP, NDF, NDM, CYCLE, AIV, APM, Ac, VA, and PG populations and three agricultural years 2017/2018, 2018/2019, and 2019/2020. In the output layer, the number of neurons corresponded to the number of defined classes, that is, 11, and the hidden layers had 64 and 128 neurons, respectively. The dataset had 7,287 examples. The algorithm used to train ANN was stochastic backpropagation (stochastic gradient descent). Adam optimizer was used. The number of training cycles was set as 600 epochs to prevent training from becoming excessive, which could lead to loss of generalization power. The ANN architecture evaluation was based on the evaluation metrics for classifiers that are mostly derived from the confusion matrix generated by the Scikit-learn package. The matrix was obtained from the test data and used to analyze the quality of predictions of the models. The other metrics used were accuracy (hit rates for positive and negative examples), precision (hit rate for positive examples), recall (coverage of correct positive examples), and F1-score (balance between

TABLE 1 | Genealogy and relative maturity group (RMG) of 11 soybean populations used in study.

Population	Genealogy	RMG
Brazil	BRS 278 RR × 5953 RSF RR	9.4/5.0
Southern	BMX Potência RR × BMX Energia RR	6.7/5.3
Northern	BRS 245 RR × BRS 278 RR	7.3/9.4
GBN1	BRS 278 RR	9.4
GB2	5953 RSF RR	5.0
GS1	BMX Potência RR	6.7
GS2	BMX Energia RR	5.3
GN2	BRS 245 RR	7.3
TGM7	TMG 1174 RR	7.4
TGM6	TMG 7262 RR	6.2
TGM8	TMG 1179 RR	7.9

precision and recall metrics). Two procedures were implemented to validate the ANN architecture. The first was the hold-out procedure, which divided the dataset into two random bases: one for the training set with 80% of the data, and the other for testing with 20% of the data. The second procedure was the k -fold cross-validation, which divided the dataset into k partitions, where $k-1$ were the data for training and k was the set used for model testing. Thus, k -models were created, where the data for training and testing were changed for each iteration (Shalev-Shwartz and Ben-David, 2014). The final model evaluation was the average of the metrics of the k models. This procedure is frequently used to validate models with relatively small datasets. k was selected as 10. For the activation of neurons, after evaluating other options, the sigmoid–logistic function was used in the hidden layers and the softmax function was applied in the output layer. The best network architecture was established based on the average accuracy, considering the evaluated possibilities, calculated by multiplying the number of neurons in each layer and the possible activation functions. Thus, the most efficient network was chosen for each strategy, adopting the lowest apparent error rate (AER) as a criterion.

RESULTS

AER and Model Evaluation Metrics

Table 2 presents the classifications of the evaluated soybean genotypes based on the 11 populations considered in the analyses conducted by the Fisher and Anderson discriminant analyses, in addition to the ANN hold-out and k -fold approaches. According to Fisher's analysis, of the total of 1,517 classifications, a total of 889 were considered as erroneous classifications, which makes an AER of 58.6%. For Anderson's methodology, in turn, for a total of 1,517 classifications, the number of erroneous classifications was 769, leading to an error rate of 50.59%, which is lower than the AER observed in Fisher. Despite this, both methods presented AER above 50%. In turn, it was observed that the validations of the ANN analyses indicated that of the 1,458 classifications performed, only 119 were considered erroneous for the hold-out approach, leading to an AER of only 8.16%, while, of the 729 classifications for the k -fold approach, only 41 were considered erroneous, constituting an even lower AER of only 5.62%.

The superiority of the k -fold cross-validation to the hold-out procedure can also be observed from **Table 3**. This table presents the metrics that assess the quality of the model used in the classification of soybean genotypes belonging to the 11 populations using the k -fold and hold-out ANN approaches. k -Fold showed accuracy of 93.36%, precision of 93.49%, recall of 93.23%, and $F1$ -score of 93.36%, which was greater than the corresponding hold-out scores by at least 1.30% for all metrics. In turn, the highest loss value was presented by the hold-out model (34.10%) when compared to the k -fold model (26.39%), in a difference of 7.71% between the two approaches.

Confusion Matrices

Considering the best parametric and nonparametric approaches, the confusion matrices generated from the validation dataset based on the classification by the Anderson's discriminant analysis and the k -fold cross-validation, are presented in **Tables 4** and **5**, respectively. To interpret a classification within a confusion matrix, the column population is that to which the genotype belongs and the row population is that allocated by the model. Therefore, the correct classifications are on the highlighted diagonal and the incorrect ones outside of it.

Table 4 shows that the allocation of the genotypes from the GS1 population in the Southern population has the largest error. The Brazil population was the only one to receive incorrect classifications from all other ten populations. Of the 23 incorrectly classified plants belonging to the Brazil population, 22 were from the Southern population. Approximately 34.5% of the plants from the Southern population and 32.6% from the population were classified as being from the Brazil population. There were no misclassifications between the Southern and northern populations.

From **Table 5**, a large reduction in the number of misclassifications can be observed. The misclassifications that occurred from the genotypes belonging to the Southern population being allocated to the Brazil population contributed most of the 5.62% of errors in the k -fold cross-validation approach. The reciprocal case of the genotypes belonging to the Brazil population being allocated to the southern population was also highlighted by the presence of errors. In addition to these pairs of populations, classification errors occurred between the Brazil and southern; GS1 and southern; southern, TGM6 and the reciprocal; and TGM6, Brazil and the reciprocal populations. The GBN1 population did not receive erroneous allocations from any other population and its genotypes were not classified as belonging to other populations.

In addition to the reduction in Anderson's errors for the k -fold cross-validation, differences in the classifications by the

TABLE 2 | Classification of soybean genotypes into 11 populations of different relative maturity groups and estimation of apparent error rate (AER) according to Fisher's and Anderson's discriminant analysis and hold-out and k -fold ANN approaches.

Approach	Total ratings	Misclassifications	AER (%)
Fisher	1,517	889	58.60
Anderson	1,517	769	50.59
Hold-out	1,458	119	8.16
k -fold	729	41	5.62

TABLE 3 | Model prediction quality evaluation metrics for hold-out and k -fold approaches in classifying soybean genotypes in 11 populations from different relative maturity groups.

Approach	Loss (%)	Accuracy (%)	Precision (%)	Recall (%)	$f1$ -score (%)
Hold-out	34.10	91.84	92.14	91.70	91.94
k -fold	26.39	93.36	93.49	93.23	93.36

two procedures were observed. In the latter, in addition to the GBN1 population, the GS2, GN2, and TGM8 populations did not receive any incorrect genotypes. Moreover, the genotypes of the Northern, GS2, TGM7, and TGM8 populations were correctly classified.

Annual Evaluation graphic

Figure 1 shows the values of incorrect classifications that occurred in the agricultural years of evaluation of the genotypes 2017/2018, 2018/2019, and 2019/2020 using the *k*-fold approach. From the increase in the bars of the graph that indicate the number of incorrect classifications, an increase in classification errors can be observed from the year 2018 to 2019 and 2019 to 2020, reaching a total increase of approximately 40%, considering the interval between the first and last years of assessment.

DISCUSSION

The failure to discriminate between genotypes and classify them into their populations correctly is caused by four main factors. First, the populations can be very similar in their origin and

genealogy. Second, the number of evaluated variables may be insufficient, in addition to having low discriminatory quality, as a third factor. The fourth cause is the use of an inadequate statistical approach (Cruz et al., 2014). In quadratic discriminant functions such as Anderson's, with an increase in the heterogeneity of variance and covariance matrices, the nonlinearity of the classification thresholds increases, enhancing the performance in modeling the structure of the function (Carvalho, 2019). The results observed in the present study corroborate with Carvalho (2019), considering that Anderson's methodology was 8.01% more accurate than Fisher's methodology. However, both discriminant functions consider parameters and assumptions that are frequently insufficient for explaining a dataset. The occurrence of the AER above 50% of the examples suggests the limitation of these methodologies in discriminating the genotypes in this study, particularly when considering the population genetic structure and their inefficiency in classifying the genotypes. In contrast, ANNs learn with experience by exploring the features contained in the data, which is capable of increasing the accuracy of the information obtained in a more detailed manner (Silva, 2019). This study proves that *k*-fold cross-validation is considered an accurate and suitable approach for small datasets (Shalev-Shwartz and Ben-David, 2014). The evaluation metrics of the models' accuracy, precision, recall, and F1-score point to the higher quality of the models as their values are closer

TABLE 4 | Classification of soybean genotypes in 11 populations from different relative maturity groups according to Anderson's discriminant analysis.

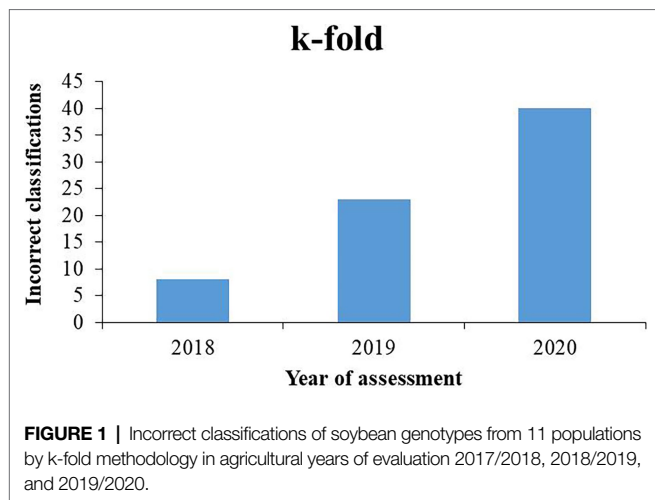
POP	Brazil	Southern	Northern	GBN1	GB2	GS1	GS2	GN2	TGM7	TGM6	TGM8
Brazil	270	96	84	2	10	70	7	15	27	28	25
Southern	22	160	0	0	19	197	3	0	0	40	12
Northern	0	0	174	3	0	0	0	15	1	0	0
GBN1	0	0	0	27	0	0	0	0	0	0	0
GB2	0	2	0	0	19	0	12	0	0	2	0
GS1	0	1	0	0	0	16	0	0	0	2	1
GS2	0	1	0	0	1	0	17	0	0	0	0
GN2	0	0	1	0	0	0	0	10	0	0	1
TGM7	0	0	1	0	0	0	0	3	20	0	7
TGM6	1	23	0	0	1	18	1	0	0	14	1
TGM8	0	0	0	0	0	1	0	0	8	4	21

The column population is that to which the genotype belongs and the row population is that allocated by the model. Therefore, the correct classifications are on the highlighted diagonal and the incorrect ones outside of it.

TABLE 5 | Classification of soybean genotypes in 11 populations of different relative maturity groups by *k*-fold approach.

POP	Brazil	Southern	Northern	GBN1	GB2	GS1	GS2	GN2	TGM7	TGM6	TGM8
Brazil	314	11	0	0	0	0	0	0	0	3	0
Southern	4	205	0	0	1	4	0	0	0	3	0
Northern	1	0	72	0	0	0	0	1	0	0	0
GBN1	0	0	0	12	0	0	0	0	0	0	0
GB2	0	2	0	0	18	0	0	0	0	0	0
GS1	0	2	0	0	0	11	0	0	0	1	0
GS2	0	0	0	0	0	0	9	0	0	0	0
GN2	0	0	0	0	0	0	0	2	0	0	0
TGM7	1	0	0	0	0	0	0	0	13	0	0
TGM6	3	4	0	0	0	0	0	0	0	25	0
TGM8	0	0	0	0	0	0	0	0	0	0	7

The column population is that to which the genotype belongs and the row population is that allocated by the model. Therefore, the correct classifications are on the highlighted diagonal and the incorrect ones outside of it.



to 100%. In turn, the loss points to a higher quality depending on how low is their presented value. The efficiency of the *k*-fold approach in correctly classifying more than 94% of the examples shows its superiority to the hold-out procedure. Moreover, the former has higher accuracy (93.36%), precision (93.49%), recall (93.23%), and f1-score (93.36%) than the latter. The higher loss presented by the hold-out approach (34.10%) compared to *k*-fold (26.39%) confirms the latter's higher classification efficiency. In terms of percentage, the less efficient ANN validation approach correctly classified 40% more genotypes than the more efficient discriminant function.

The proximity of the origin and genealogy of genotypes belonging to different populations largely contributes to the inefficiency of their discrimination, as also observed by Sant'anna (2014) in the study on ANN and backcross populations with different degrees of similarity. Anderson's discriminant function was inefficient in discriminating genitors from their descendants because most of the incorrect classifications that occurred with GS1 genotypes, represented by the BMX Potência RR cultivar, were allocated to the southern population, which is its descendant.

The Brazil population presents a high level of genetic variability in the relative maturity, resulting in the presence of many allelic combinations of genes that control the timing of flowering and maturity (Lin et al., 2021). This was also a limiting factor for the good performance of the parametric methodology, causing it to classify genotypes from all other populations as belonging to the Brazil population, equivalent to 24% of the total errors. This capitalizes all variations in populations with narrow genetic basis and cultivars of pure lineages. Despite this, in the Brazil population, only 23 plants were misclassified, that is, 7.8%, and 22 of these were allocated to the Southern population. This shows that it is broadly representative phenotypically of all other populations, even with little similarity to the Southern population, characterizing the wide genetic variability in the Brazil population. In contrast, 34.5% of the plants in the Southern population were classified as being from the Brazil population, which does not present a relationship or known biological explanation. Thus, the applied statistical technique is the probable cause of the incorrect classifications (Cruz et al., 2014). In the Northern population,

32.6% of the individuals were classified into the Brazil population. In this case, the two populations have a common genitor, which may also be a biological cause of the misclassification. Between the Northern and Southern populations, there were no misclassifications, and there were also no relationships between the genitors.

When considering the genitors of the populations and the other three witnesses, it was noted that in the Northern population, fewer misclassifications occurred compared with that in the other populations; however, the errors of classification, in general, were always large, both in absolute values and percentages. There was a tendency for no differentiation of the genitor with the derived population, as in the case of GS1, Brazil and Southern populations and GN2, Brazil and Northern populations. Witnesses, even being pure lines, were erroneously classified as belonging to different populations, mainly for plants of the Southern population in the TGM6 witness. Thus, the statistical technique is inefficient in classifying pure lineage genotypes or those with a narrow genetic base and is partially effective even in populations with a broad genetic base, such as the Brazil population.

The only case in which no error occurred using Anderson's methodology was the GBN1 population, corresponding to the BRS 278 RR cultivar with RMG 9.4, which did not receive incorrect genotypes from any other population. The RMG of this population, considered late for the evaluation region, induces elongation of the vegetative stage, resulting in extremely tall plants with low grain production, contributing to an atypical phenotype of this cultivar in Jaboticabal, the ideal RMGs for which are between 6 and 8.

Based on the confusion matrix obtained by the *k*-fold cross-validation, the wide variability of the Brazil population also led to misclassifications of the genotypes from the Southern population. However, the number of errors by this method was 11 compared to 96 errors by Anderson's discriminant analysis for the above pair of populations. The other *k*-fold misclassifications comprising the 5.62% errors occurred between Brazil and southern; GS1 and Southern; southern, TGM6 and the reciprocal; and TGM6, Brazil and the reciprocal populations. However, the number of errors was always equal to or less than 4. The non-occurrence of misclassification errors in the GBN1, GS2, GN2, and TGM8 populations as well as the absence of incorrect allocations in the Northern, GBN1, GS2, TGM7, and TGM8 populations demonstrate that, even with a relationship between the genitor and its derived population, ANN is very efficient for broad-based populations. Furthermore, the ANN correctly classifies the narrow-based populations and pure lines, learning to differentiate genotypes. Although the classification scheme of the ANN was affected by the genotype, it was to a lesser extent than Anderson's discriminant method.

The mechanical evaluations based on the traditional characteristics in the three considered years at the same location did not impair the performance of the ANN; however, it shows the poor influence of noise and data loss on the quality of the information obtained (Silva, 2019). Furthermore, the correct identification of the genitors and witnesses shows that the genetic proximity between the genotypes and the existing genetic variability, whether wide or restricted, does not limit to the

result accuracy. This demonstrates the usefulness of the technique in breeding programs even in the F_6 and F_7 generations.

The built ANN identified the genotypes that were incorrectly classified by their corresponding numbers in the data table. From this information, it was possible to identify the generation of self-fertilization of each of these genotypes based on the year of evaluation. As the generations advance in homozygosity and selections are made within the populations, mainly based on productivity, they become uniform and similar to each other because they are generally intended for cultivation where they are located, as also identified by Sant'anna (2014) using simulated data. Genotypes grown outside their ideal photoperiod conditions present reduced productive potential (Lin et al., 2021) and are eliminated in the selection process. The increase in incorrect classifications from 2018 to 2020 reached approximately 40%, which shows the importance of classifying and assigning genotypes to their appropriate evaluation regions. At this stage of the local breeding program, it would be ideal if the advanced lines in F_5 or F_6 were evaluated in the target region according to their photoperiod. The dataset also leads to the conclusion that lineage development in an intermediate region (RMG 7) is suitable for advancing the lineages for extreme photoperiods, as for RMGs 5 and 9.

In a conventional soybean breeding program, many resources are used to obtain a successful cultivar. Time and money are invested in researching, testing, selecting, and developing promising genotypes. After the selection stage, it takes at least 5 years for the implementation of tests in the intended cultivation regions for the commercialization of the cultivar and it can be extended up to 10 years until the launch of the final product (commercial cultivar) (Seednews, 2019). This study demonstrated that neural networks are able to classify genotypes and discriminate them for their ideal ranges of cultivation even in early generations, which would lead to a reduction of resources spent between the selection and testing steps in the existing edaphoclimatic macroregions. In addition, the high efficiency of ANNs in discriminating populations with wide and narrow genetic variability allows their application to obtain genotypes to be tested throughout Brazil, from the same base population of wide genetic diversity, even in early and intermediate generations of inbreeding, for the development of new cultivars, enabling the exploration of greater genetic variability from wider crosses and contributing once again to the reduction of time and resources invested in genetic improvement programs.

REFERENCES

- Alliprandini, L. F., Abatti, C., Bertagnolli, P. F., Cavassim, J. E., Gabe, H. L., Kurek, A., et al. (2009). Understanding soybean maturity groups in Brazil: environment, cultivar classification, and stability. *Crop Sci.* 49, 801–808. doi: 10.2135/cropsci2008.07.0390
- Carvalho, V. P. (2019). Aprendizado de máquina e estatístico na discriminação de populações na presença de matrizes de covariâncias heterogêneas e vetores aleatórios não normais multivariados. [These]. Viçosa (MG): Universidade Federal de Viçosa
- Cruz, C. D. (2008). *Programa genes: diversidade genética*. Viçosa: Universidade Federal de Viçosa.
- Cruz, C. D., Regazzi, A. J., and Carneiro, P. C. S. (2014). *Modelos biométricos aplicados ao melhoramento genético*. Viçosa: Universidade Federal de Viçosa.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

LA: planning, implementation, and experimental conduction, field evaluation, analysis and data collection, and text writing. GM: analysis and data collection, text writing, and correction and guidance. BV: implementation and experimental conduction, and analysis and data collection. AS: implementation and experimental conduction, field evaluation, and analysis and data collection. AM: implementation and experimental conduction, field evaluation, and analysis and data collection. SU-T: planning, analysis and data collection, text writing; correction and guidance, and work leader. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Pro-rectory of Postgraduation (PROPG), São Paulo State University - UNESP/FCAV Jaboticabal.

ACKNOWLEDGMENTS

We thank the State University “Júlio de Mesquita Filho”—Unesp/FCAV Jaboticabal Campus and the Graduate Agronomy Program (Genetics and Plant Breeding) and the Coordination for the Improvement of Higher Education Personnel (CAPES).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.814046/full#supplementary-material>

Supplementary Table | Dataset collected from soybean progenies belonging to different relative maturity groups, during three agricultural years of evaluation.

- do Carmo Oda, M., Sediya, T., Matsuo, E., Nascimento, M., and Cruz, C. D. (2019). Estabilidade e adaptabilidade de produção de grãos de soja por meio de metodologias tradicionais e redes neurais artificiais. *Sci. Agr. Paran.* 18, 117–124.
- Federer, W. T. (1956). Augmented (hoonuiaku) designs. *Hawaiian Planters' Record, Aica* 55, 191–208.
- Garner, W. W., and Allard, H. A. (1930). Photoperiodic response of soybeans in relation to temperature and other environmental factors. *J. Agric. Res.* 41, 719–735.
- Hang, J., Zhang, D., Chen, P., Zhang, J., and Wang, B. (2019). Classification of plant leaf diseases based on improved convolutional neural network. *Sensors (Basel)*. 19:4161. doi: 10.3390/s19194161
- Haykin, S. (2008). *Neural Networks and Learning Machines*. (3rd Edn.). Pearson-Prentice Hall, Hamilton.

- Johnson, R. A., and Wichern, D. W. (2002). *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall.
- Kavzoglu, T., and Mather, P. (2003). The use of backpropagation artificial neural networks in land cover classification. *Int. J. Remote Sens.* 24, 4907–4938. doi: 10.1080/0143116031000114851
- Lin, X., Liu, B., Weller, J. L., Abe, J., and Kong, F. (2021). Molecular mechanisms for the photoperiodic regulation of flowering in soybean. *J. Integr. Plant Biol.* 63, 981–994. doi: 10.1111/jipb.13021
- Lu, W., du, R., Niu, P., Xing, G., Luo, H., Deng, Y., et al. (2022). Soybean yield Preharvest prediction based on bean pods and leaves image recognition using deep learning neural network combined with GRNN. *Front. Plant Sci.* 12:791256. doi: 10.3389/fpls.2021.791256
- Miladinović, J., Čeran, M., Đorđević, V., Balešević-Tubić, S., Petrović, K., Đukić, V., et al. (2018). Allelic variation and distribution of the major maturity genes in different soybean collections. *Front. in plant Sci.* 9:1286. doi: 10.3389/fpls.2018.01286
- Miladinović, J., and Đorđević, V. (2011). “Soybean morphology and stages of development,” in *Soybean*. eds. J. Miladinović, M. Hrustić and M. Vidić (Bečej: Sojaprotein), 45–68.
- Rahimi, Y., Bihamta, M. R., Taleei, A., Alipour, H., and Ingvarsson, P. K. (2019). Applying an artificial neural network approach for drought tolerance screening among Iranian wheat landraces and cultivars grown under well-watered and rain-fed conditions. *Acta Phys. Plant.* 41:156. doi: 10.1007/s11738-019-2946-2
- Sá, L.G. (2018). Inteligência computacional aplicada ao estudo da divergência e fenotipagem em cultivares de soja. dissertation. Montes Claros (MG): Universidade Federal de Minas Gerais.
- Samanfar, B., Molnar, S. J., Charette, M., Schoenrock, A., Dehne, F., Golshani, A., et al. (2017). Mapping and identification of a potential candidate gene for a novel maturity locus, E10, in soybean. *Theor. Appl. Genet.* 130, 377–390. doi: 10.1007/s00122-016-2819-7
- Sant’anna, I. C. (2014). *Redes neurais artificiais na discriminação de populações de retrocruzamento com diferentes graus de similaridade*. dissertation. Viçosa: Universidade Federal de Viçosa.
- Seednews (2019). Os desafios na criação de novas cultivares. Available at: <https://seednews.com.br/artigos/3080-os-desafios-na-criacao-de-novas-cultivares-edicao-novembro-2019>. Accessed March 25, 2022.
- Shalev-Shwartz, S., and Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. Cambridge: Cambridge University Press.
- Silva, W. D. M. (2019). Redes neurais artificiais como ferramenta para prognose de crescimento e melhoramento genético florestal. [These]. Jaboticabal: Universidade Estadual Paulista “Júlio de Mesquita Filho”.
- Soares, F. C., Robaina, A. D., Peiter, M. X., and Russi, J. L. (2015). Predição da produtividade da cultura do milho utilizando rede neural artificial. *Ciência Rural* 45, 1987–1993. doi: 10.1590/0103-8478cr20141524
- Sudheer, K. P., Gosain, A. K., and Ramasastri, K. S. (2003). Estimating actual evapotranspiration from limited climatic data using neural computing technique. *J. Irrig. Drain. Eng.* 129, 214–218. doi: 10.1061/(ASCE)0733-9437(2003)129:3(214)
- Taratuhin, O. D., Novikova, L. Y., Seferova, I. V., Gerasimova, T. V., Nuzhdin, S. V., Samsonova, M. G., et al. (2020). An artificial neural network model to predict the phenology of early-maturing soybean varieties from climatic factors. *Biophysics* 65, 106–117. doi: 10.1134/S0006350920010200
- Trivedi, N. K., Gautam, V., Anand, A., Aljahdali, H. M., Villar, S. G., Anand, D., et al. (2021). Early detection and classification of tomato leaf disease using high-performance deep neural network. *Sensors (Basel)*. 21:7987. doi: 10.3390/s21237987
- 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 Amaral, Miranda, Val, Silva, Moitinho and Unêda-Trevisoli. 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.



Exploring Soybean Flower and Pod Variation Patterns During Reproductive Period Based on Fusion Deep Learning

Rongsheng Zhu^{1*}, Xueying Wang², Zhuangzhuang Yan², Yinglin Qiao², Huilin Tian³, Zhenbang Hu³, Zhanguo Zhang¹, Yang Li¹, Hongjie Zhao¹, Dawei Xin³ and Qingshan Chen^{3*}

¹ College of Arts and Sciences, Northeast Agricultural University, Harbin, China, ² College of Engineering, Northeast Agricultural University, Harbin, China, ³ College of Agriculture, Northeast Agricultural University, Harbin, China

OPEN ACCESS

Edited by:

Deyue Yu,
Nanjing Agricultural University, China

Reviewed by:

Renata Retkute,
University of Cambridge,
United Kingdom
Karansher Singh Sandhu,
Bayer Crop Science, United States

*Correspondence:

Rongsheng Zhu
rshzhu@126.com
Qingshan Chen
qshchen@126.com

Specialty section:

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

Received: 17 April 2022

Accepted: 20 June 2022

Published: 13 July 2022

Citation:

Zhu R, Wang X, Yan Z, Qiao Y,
Tian H, Hu Z, Zhang Z, Li Y, Zhao H,
Xin D and Chen Q (2022) Exploring
Soybean Flower and Pod Variation
Patterns During Reproductive Period
Based on Fusion Deep Learning.
Front. Plant Sci. 13:922030.
doi: 10.3389/fpls.2022.922030

The soybean flower and the pod drop are important factors in soybean yield, and the use of computer vision techniques to obtain the phenotypes of flowers and pods in bulk, as well as in a quick and accurate manner, is a key aspect of the study of the soybean flower and pod drop rate (PDR). This paper compared a variety of deep learning algorithms for identifying and counting soybean flowers and pods, and found that the Faster R-CNN model had the best performance. Furthermore, the Faster R-CNN model was further improved and optimized based on the characteristics of soybean flowers and pods. The accuracy of the final model for identifying flowers and pods was increased to 94.36 and 91%, respectively. Afterward, a fusion model for soybean flower and pod recognition and counting was proposed based on the Faster R-CNN model, where the coefficient of determination R^2 between counts of soybean flowers and pods by the fusion model and manual counts reached 0.965 and 0.98, respectively. The above results show that the fusion model is a robust recognition and counting algorithm that can reduce labor intensity and improve efficiency. Its application will greatly facilitate the study of the variable patterns of soybean flowers and pods during the reproductive period. Finally, based on the fusion model, we explored the variable patterns of soybean flowers and pods during the reproductive period, the spatial distribution patterns of soybean flowers and pods, and soybean flower and pod drop patterns.

Keywords: soybean, fusion model, flower, pod, deep learning

INTRODUCTION

Soybeans are an important food crop all around the world, and are also a major source of oil and protein (Singh et al., 2010). In recent years, many scholars have worked on yield components, such as the number of grains and pods per plant in order to improve soybean yield (Li et al., 2018, 2019; Du et al., 2019; Song et al., 2020). However, the attempts made so far have not fundamentally improved yields (Zhang et al., 2012), so it is essential to find the main factors affecting soybean yields in order to increase them. Soybean flower and pod drops are a common phenomenon during the growth and development of soybeans. It occurs from the emergence of flowers buds

to flowering and all stages of podding. The flower and pod drop rates can even reach 30 to 80%. (Selvaraj et al., 2019). Therefore, reducing the flower and pod drop can significantly increase soybean yield (Wang et al., 2010).

In the study of flower and pod drop patterns, many researchers have explored different aspects of the flower and pod patterns of soybeans over the years. Gao et al. (1958) first investigated the order of flower abscission in soybeans and concluded that the order of flower abscission is identical to the order of flowering, with early flower abscission being higher than late flower abscission. Ma et al. (1960) observed average flower drop, pod drop, and bud drop rates of 56.5, 28.4, and 13.1%, respectively, from 163 soybean plants. Song and Dong (2002); Su et al. (2004) conducted an in-depth study on the flowering order, and concluded that the flowering order of limited pod habit soybeans starts from the middle and gradually opens upward and downward, while the flowering order of both sub-limited and unlimited pod habit soybeans opens sequentially from the bottom upward. Zhang et al. (2010) localized the QTL for the number of flowers and pods per plant trait from a genetic perspective and identified gene regions associated with flowering and pods setting. Zhao et al. (2013) broke away from the conventional approach of conducting static observations of flowering and flower drop statistics, and studied flowering patterns from both spatial and temporal perspectives. They concluded that the number of flowers per day varied significantly among different soybean flowering stages, with different flowering nodes and durations. Meanwhile, the drop rate of flowers at different stages also differed significantly. However, there is a gap in this area of research abroad, and, in the studies mentioned above, the indicators were done manually, which can be time-consuming and labor-intensive, and the surveys are subjective, difficult, and error-prone (Bock et al., 2010; Singh et al., 2021). Because of these drawbacks, it is impractical to carry out flower and pod phenotypic surveys manually for large fields and difficult to implement on large samples, thus preventing more general conclusions from being drawn. On the other hand, the complex structure of the soybean plant and the severe shading during the growing season have made many fine phenotypes unavailable, and even important phenomena, such as flower and pod abscission, have been stalled by the difficulty of phenotypic investigations and the lack of real-time, accurate, and bulk phenotypic support for their patterns.

In recent years, there have been significant advances in computer vision technology, largely thanks to the development of neural network techniques, such as deep learning (Lecun et al., 2015). Singh et al. (2021) proposed machine vision as a method to address duress severity phenotyping, which could improve the speed, accuracy, reliability, and scalability of image-based disease phenotyping. Deep learning is a kind of the machine learning method, which predicts complex and uncertain phenomena by learning a large amount of data. DL has led to the pattern transformation of the image-based plant phenotype. This method is not only used for the digital image-based plant adversity phenotype, but also performs well in a wide range of plant phenotype tasks, such as leaf counting (Ubbens et al., 2018), flowering detection (Xu et al., 2018),

and plant identification (Šulc and Matas, 2017) with good results (Singh et al., 2018). Gill et al. also introduced the application of machine learning and deep learning methods in plant stress phenotypes. The DL models used for Phenomics mainly include multilayer perceptron, generative antagonism network, convolutional neural network (CNN), and recurrent neural network. CNN has great advantages in image analysis, and different CNN networks are used for different plants (Gill et al., 2022). CNNs in deep learning demonstrate powerful feature extraction capabilities for images and are widely used in image-based agricultural computer vision tasks. In addition, deep learning also plays an important role in field phenotype counting. Many scholars have used deep learning techniques to identify and count different research targets. Rahneemoonfar and Sheppard (2017) proposed a method for estimating fruit counts based on a deep CNN regression model using synthetic data for the training of the network, achieving 91% accuracy on a real data set, which is robust under adverse conditions. Lu et al. (2017) proposed the TasselNet model for counting maize males in complex field environments to estimate them as an output density map. TasselNet has greatly reduced counting errors, but field counting of maize ears is still an open question. Xiong et al. (2019) improved TasselNet by using context-enhanced context-linked local regression networks for field counts of wheat ears, further enhancing the accuracy of TasselNet for field ears by up to 91.01% on the dataset. Hasan et al. (2018) proposed a wheat-counting algorithm based on the Faster R-CNN's object detection model, but not tested at higher densities. Ghosal et al. (2019) proposed different object detection algorithms for high accuracy detection of sorghum heads in UAV images, all with an accuracy of around 95% and robustness to changing directions, as well as different lighting conditions. Wu et al. (2019) proposed a rice grain-counting method based on Faster R-CNN object detection, with an accuracy higher than 99%, which has a very high accuracy for calculating the number of rice grains per spike. Mu et al. (2020) applied an object detection model to the identification and counting of unripe and ripe tomato fruits at all periods, with an accuracy of 87.83% and good detection accuracy for highly shaded unripe tomatoes in real cultivation scenarios. David et al. (2020) applied the object detection model for experiments on the detection of different periods and varieties of wheat ears, and the results showed that the object detection model has good accuracy, stability, and robustness. Riera et al. (2020) also proposed a combination of object detection and RGB images that could count pods in the R8 period, and thus make a soybean yield prediction. However, they only counted pods in the R8 period, in contrast to our work, where we counted pods from the beginning pod stage to the full maturity stage (R3 to R8). The above methods based on CNNs and regression models focus more on the number of objects present in the image and do not give the position of the objects in the image, and are often used for automatic counting in high-density, high-volume fields; deep networks based on object detection can not only automatically count objects in images but also locate and track them. This shows that the object detection class of models based on deep learning algorithms is more suitable for the localization and identification of soybean flowers and pods.

In this paper, we optimized the Faster R-CNN model for the characteristics of soybean flowers and pods, respectively, and proposed a fusion model based on the Faster R-CNN two-stage object detection algorithm to identify and count flower and pod simultaneously. The fusion model was then used to explore and analyze three aspects of soybean fertility: patterns of flower and pod variation, patterns of flower and pod spatial distribution, and patterns of flower and pod drop in soybeans.

MATERIALS AND METHODS

Experimental Materials

In this paper, four soybean cultivars, namely, DongNong252 (DN252), HeiNong51 (HN51), ZheNong No. 6, and ChunFengZao, were selected as experimental soybean flower samples in three fields in Harbin, Heilongjiang province, Fuyang, Anhui province, and Hangzhou, Zhejiang province, China. The Harbin experimental field in Heilongjiang province is located at an east longitude of $125^{\circ}42' - 130^{\circ}10'$ and north latitude of $44^{\circ}04' - 46^{\circ}40'$, the Fuyang experimental field in Anhui province is located at an east longitude of $114^{\circ}52' - 116^{\circ}49'$ and north latitude of $32^{\circ}25' - 34^{\circ}04'$, the Hangzhou experimental field in Zhejiang province is located at an east longitude of $118^{\circ}21' - 120^{\circ}30'$ and north latitude of $29^{\circ}11' - 30^{\circ}33'$. Among them, DN252 and HN51 have a sub-limited pod habit, while ChunFengZao and ZheNong No. 6 have a limited pods habit. The basic information on image acquisition is shown in **Table 1**. A total of 1,895 images ($3,024 \times 4,032$ pixels) were captured. The soybean pod test samples were planted at the Northeast Agricultural University experimental field base, and the two cultivars tested were DN252 and HN51. A total of 2,693 images ($3,024 \times 4,032$ pixels) were acquired. Two soybean cultivars, DongNong252 (DN252) and HeiNong51 (HN51), which are mainly grown in Harbin, Heilongjiang Province, were selected as the test samples to study the variation pattern.

In our data samples, the samples with red spider disease are included. Red spider disease damages soybean leaves. At the early stage of damage, yellow and white spots appear on the front of

soybean leaves. After 3–5 days, the spot area expands and the spots are dense, and the leaves begin to appear reddish brown patches. With the aggravation of the damage, the leaves turn rusty brown, curl, and, finally, fall off. When spider disease is serious, it will affect the size of soybean flowers and the number of pods. However, at the beginning of suffering from spider disease, we used 2,000–3,000 times of 1.8% avermectin EC to treat the diseased plants. We sprayed the medicine every 5 days, focusing on the back of the upper tender leaves, tender stems, and flower organs of the plants, and sprayed the medicine evenly. After spraying, the soybean plants have completely recovered to normal, and there is no obvious abnormality in flowers and pods, so it has no impact on our identification and counting.

Image Acquisition and Processing

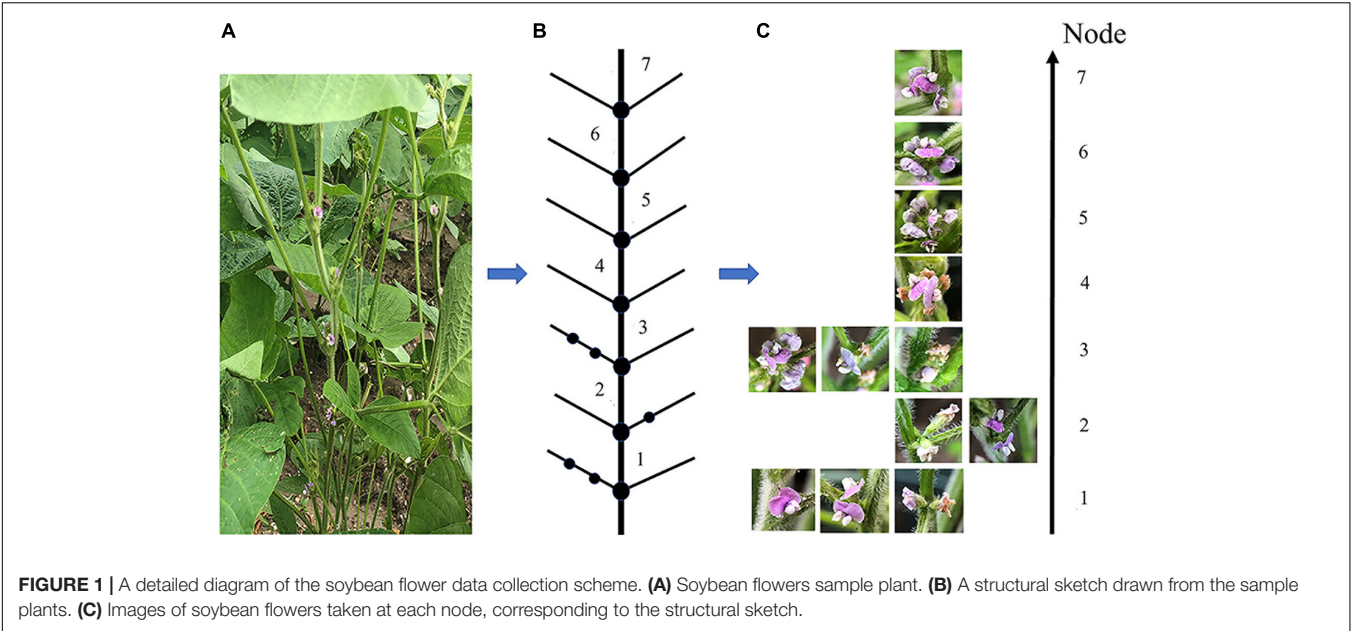
Due to soybean plants having a complex structure, various problems made it impossible to obtain a clear picture of all the soybean pods in a single image. These issues include pods and leaves being similar in color, leaves blocking pods and flowers, pods blocking one another, soybean flowers being stacked, flowers too small to see, etc. Nodal shots were used for the soybean flower and pod images. To view the soybean flower characteristics, the viewing angle was overhead, the angle between the phone and the main stem was about 25 degrees, and the distance from the node was 10 to 15 cm; the pods with unobstructed nodes were photographed from a flat angle, and those with heavily obscured nodes from a top angle. The details of the soybean flower data collection method are presented in **Figure 1**. **Figure 1A** features a photograph of the flowers at each node on the plant. **Figure 1B** is a sketch of the sample structure from the overall photograph of the sample plants, with the lowest node number of the soybean structure marked as 1, and the higher nodes numbered sequentially. As seen in **Figure 1C**, the photographed node soybean flower image corresponds to the individual nodes of the structural sketch.

Flower and pod drop pattern test samples were planted in pots on 25th May 2019 in the Northeast Agricultural University experimental field. Three plants of each variety were selected for observation, and six plants were observed daily at 8 a.m., 12 noon, and 5 p.m. Flowering began on 25th June (R1) and finished on 30th September (R8); the number of flowers per plant, the flowers at the nodes, and the flowers dropped were recorded, and the flowering nodal positions and nodal position of the flowers dropped were photographed. The plants were affected by red spider disease during planting, and, despite being promptly treated with pesticides, some of the sample plants were still affected, rendering it impossible to repeat the trial due to the effects of new crown pneumonia.

The soybean flower and pod datasets were reduced to 640×640 pixels using image processing techniques. All datasets were produced in the PASCAL VOC format, and the flowers and the pods in the images were manually labeled as real bounding boxes using the LabelImg tool (Tzutalin, 2018). After production, the datasets were randomly divided into a training set, a validation set, and a test set, with the ratio of the training set plus the validation set to test the set at 8:2. The training and validation sets are then divided in the ratio of 9:1. The specific

TABLE 1 | Basic information on image acquisition.

Variety	Image acquisition time	Podding habit	Number of images	Color of flower
DN252	2019	Sub-limited podding habit	568	White
ChunFengZao	2020	Sub-limited podding habit	545	White
ZheNong NO. 6	2020	Limited podding habit	266	Purple
HN51	2019	Limited podding habit	516	Purple



division of the soybean flower and pod datasets is presented in **Table 2**.

box. In this paper, we used both Resnet-50 and VGG16 backbone for the training of the feature extraction network.

Models

Faster R-CNN (Resnet-50) and Faster R-CNN (VGG16)

Faster R-CNN is a fast, two-stage target detection method, developed after the R-CNN and Fast R-CNN methods, which has been widely used in various fields (Rampersad, 2020). The Faster R-CNN object detection algorithm is divided into four main parts: Feature Extraction Network, Region Proposal Network (RPN), RoI Pooling Layer, and Fully Connected Layer. The network structure is displayed in **Supplementary Figure 1A**. The RPN replaces the previous Selective Search method and is used to generate candidate boxes, classify and determine whether the set anchor contains the detection target, and perform bounding box regression. RoI Pooling is used to collect the RPN-generated proposals and extract them from the feature maps in the feature extraction network, generating feature maps to be fed into the subsequent fully connected layers for further classification and regression. Finally, we use the proposed feature maps to calculate the specific category and perform another bounding-box regression to obtain the exact final position of the detection

Single Shot MultiBox Detector

Compared to the two-stage algorithm Faster R-CNN, the first proposed YOLO algorithm has a significantly faster detection speed, but its accuracy rate is inadequate. Consequently, Liu et al. (2016) proposed a Single Shot MultiBox Detector (SSD). The SSD algorithm uses multi-scale features and default anchor boxes to detect the presence of multi-scale objects in the scene in a single step (Akshatha et al., 2022). Its network structure is shown in **Supplementary Figure 1B**. The SSD network uses a one-stage network to solve the detection speed problem and adopts the anchors idea from Faster R-CNN. Feature extraction is performed in layers and border regression, and classification operations are computed sequentially, allowing training and detection tasks to be adapted to multiple target scales. The elements extracted from the shallow layer help to detect smaller objects, while the deeper layer elements are responsible for detecting larger objects. SSD networks are divided into six stages, each of which learns a feature map and then performs border regression and classification.

EfficientDet

EfficientDet, like SSD, is a single-shot detector with an anchor-based target detection method that has achieved good results in both detection speed and accuracy. EfficientNet (Tan and Le, 2019) is the backbone of the network, and the BiFPN feature network extracts the P3, P4, P5, P6, and P7 features in EfficientNet and performs a bidirectional feature fusion. BiFPN (bidirectional feature network) is the core part of the EfficientDet network for fast multi-scale feature fusion (Tan et al., 2020). The fused features are then sent to the class prediction and bounding box prediction networks, which generate the class and bounding box positions of the objects. The class and bounding box network

TABLE 2 | Classification of the data set for soybean flowers and pods.

Dataset	Division	Number	Dataset	Division	Number
Soybean flower	Training set	1,364	Soybean pod	Training set	1,938
	Validation set	152		Validation set	216
	Test set	379		Test set	539
	Total number	1,895		Total number	2,693

weights are shared between all feature levels. The EfficientDet network structure is presented in **Supplementary Figure 1C**.

YOLOV3

The popular target detection algorithm YOLOV3 (Redmon and Farhadi, 2018) has significant advantages in terms of speed and accuracy (Wang and Liu, 2020). Its backbone network is Darknet53, which uses the 52 layers in front of Darknet-53 (no fully connected layers). This network is a fully convolutional network that makes extensive use of residual hopping connections in order to reduce the negative effect of gradients from pooling, directly removes pooling, and uses the stride in conv to achieve down-sampling. The FPN-like up-sample and fusion approach is used in YOLOV3 to perform detection on multiple feature map scales, improving the detection of mAP and small objects (Redmon and Farhadi, 2018). The network structure is shown in **Supplementary Figure 1D**.

YOLOV5

YOLOV5 is the latest algorithm in the YOLO series, and it performs better than other YOLO networks in terms of accuracy and speed (Thuan, 2021). Its network structure, with two different CSPs used in Backbone and Neck, is shown in **Supplementary Figure 1E**. In Backbone, CSP1_X with a residual structure is used since the Backbone network is deeper and the addition of the residual structure enhances the gradient value when back-propagating between layers. This effectively prevents the gradient from disappearing as the network deepens, resulting in finer features. Using CSP2_X in Neck, the backbone network output is split into two branches as opposed to one simple CBL, which is later concatenated, enhancing the network's ability to fuse features and retain richer feature information. Furthermore, it uses focus for image slicing operation in Backbone and the FPN+PAN structure in Neck.

Fast-SCNN

The Fast-SCNN network consists of 4 parts: a down-sample learning module, a global feature extractor, a feature fusion module, and a classifier (Poudel et al., 2020). The network structure is shown in **Supplementary Figure 2A**, where it can be seen that the 2 branches share the down-sample learning module to further reduce the computational effort. The learning to down-sample a module consists of 3 convolutional layers, the first of which is a normal convolutional layer and the following two use depth-separable convolution to improve computational efficiency. The global feature extractor is used to extract global features and the learning to down-sample module output feature map is fed into the depth branch of the 2-branched structure. The bottleneck residual block proposed in MobileNet-v2 is used to construct the global feature extractor, where the depth-separable convolution helps to reduce both the number of parameters and the computational effort required (Sandler et al., 2018). The global feature extractor also contains a pyramid pooling module to extract contextual features at different scales (Zhao et al., 2000). The feature fusion module is used to fuse the output features of the 2 branches. Fast-SCNN uses a relatively simple structure for feature fusion to maximize computational efficiency.

The classifier module contains two deeply separable convolutions and a convolutional kernel (size, 1×1) to improve network performance. SoftMax is also included in the classifier module to calculate the training loss.

Image Cascade Network

Image Cascade Network, or ICNet (Zhao et al., 2018), is fed with multiple varying resolution images in order to balance accuracy and speed. Low resolution images are fed into an PSPNet network, called a heavy CNN, where parameters are shared between the branches of low- and medium-resolution images, thus reducing the execution time. High-resolution images are then fed into a light CNN. ICNet throws low-resolution images into a complex CNN and high-resolution images into a lightweight neural network. CFF and cascade label guidance are then used to integrate high resolution features and gradually refine the coarse, low-resolution semantic graph. The ICNet network structure is illustrated in **Supplementary Figure 2B**.

U-Net

The U-Net network structure (Navab et al., 2015) is relatively simple compared to other networks, with Encoder on the left operating as a feature extractor and Decoder on the right operating as an up-sampler. Since the overall network structure is similar to a U shape, it is known as U-Net, and its network structure is shown in **Supplementary Figure 2C**. The Encoder consists of a convolution operation and a down-sampling operation. The convolution structure is a uniform 3×3 convolution kernel with 0 padding and a striding of 1. The above two convolutions are followed by a max pooling with a stride of 2. The output size becomes $\frac{1}{2}(H, W)$. The above steps are repeated 5 times, the final time without max pooling, and the resulting feature map is fed directly into Decoder. The feature map is restored to its original resolution by Decoder, using convolution, up-sampling, and skip connection.

DeepLabV3+

The DeepLabV3+ network (Firdaus-Nawi et al., 2011) structure is shown in **Supplementary Figure 2D**, which adds simple Decoder to DeepLabV3+ to refine the segmentation results, has good target boundary detection results, and is implemented in a two-in-one spatial pyramid pool module or codec structure. In Encoder, DCNN is the backbone network for feature extraction, and Atrous Spatial Pyramid Pooling is based on SPP with Atrous Convolution, which is used for feature extraction with different rates of Atrous Convolution. The features are then compressed through being concat-merged and 1×1 convolved. In Decoder, the low-level features are dimensionally adjusted by 1×1 convolution and the high-level features are upsampled by a factor of 4 to adjust the output stride. The features are then concat, followed by 3×3 convolution, before being upsampled by a factor of 4 to obtain the output prediction.

Establishment of a Fusion Model

Due to the complex soybean structure, a highly accurate and stable phenotype extraction model is necessary for the process of flower and pod phenotype extraction. Firstly, we selected some

classical models to simultaneously identify soybean flowers and pods, and discovered that the Faster R-CNN algorithm object detection was the most accurate, much greater than any other recognition and segmentation models, but there is still room for improvement, so the Faster R-CNN algorithm was fine tuned and a fusion model was finally proposed.

Faster R-CNN is a classical, two-stage, object detection algorithm based on CNNs, which has been widely used in several fields since it was proposed. The model contains four main components: a feature extraction network, a RPN, an RoI pooling layer, and a fully connected layer. The feature extraction network, also known as the backbone network, is generally composed of a CNN that extracts features from the input image to obtain a feature map for use in the proposal module and RoI pooling. Studies have shown that the backbone feature extraction ability largely determines the network performance, and a large deep backbone can effectively improve detection. Therefore, generally speaking, the deeper the network, the better the results, and the more difficult it is to train. However, as the depth increases, so does the network training cost and the effect decreases instead of increasing (He et al., 2016).

The benchmark Faster R-CNN uses a pre-trained VGG16 CNN as the feature extraction network. In the targeting soybean flower recognition and counting section, we used ResNet-50, pre-trained on the ImageNet dataset (Fei-Fei et al., 2010), instead of VGG16 as a feature extraction network to extract deeper flower features as the ResNet-50 residual structure enables the fusion of shallow and deep features, solving the gradient disappearance problem and making the model easier to train.

The backbone selection is particularly critical to soybean pod identification and counting, and four backbone networks were employed: ResNet-50, ResNet-101, CSPResNet-50, and CSPResNet-101, in combination with Feature Pyramid Network (FPN) modules (Lin et al., 2017) for soybean pod identification, whereas Composite Backbone Network was proposed by Liu et al. (2019). The model is composed of different, or multiple, backbones, in a phase-by-phase iterative manner using the output features of the previous backbone as part of the input features for the subsequent backbone, and, finally, using the feature mapping of the last backbone for object detection. The FPN module can fuse lower-level features with less feature semantic information, and higher-level features with more semantic information, to independently perform predictions at different feature levels. Analysis of the results resulted in CSPResNet-50 combined with FPN being used as our pod recognition and counting feature extraction network, where ResNet-50 was chosen as the basic network for the CSPResNet-50 network.

Secondly, according to the characteristics of the large difference in pod size over different periods, we analyzed the pixel area occupied by the 6,334 bounding boxes annotated in the training set, as shown in **Figure 2**. **Figure 2A** shows the distribution of the number of bounding box areas, and **Figure 2B** shows the percentage of the area of the bounding box in the image.

Following the same criteria as the Microsoft COCO dataset (Lin et al., 2014), we took the number of pixels occupied by the

bounding box as the area of each pod instance, divided into small, medium, and large scales, according to the following conditions:

1. Small target: the area of pixels occupied by the bounding box $\leq 32 \times 32$;
2. Medium target: $32 \times 32 \leq$ the area of pixels occupied by the bounding box $\leq 96 \times 96$;
3. Large target: the area of pixels occupied by the bounding box $\geq 96 \times 96$.

Taken together, the percentage of annotated bounding boxes in the image ranges from 0 to 0.25, with more significant scale variation, and a large proportion of the annotated bounding boxes are within the small target range. These small targets are likely to be pods at R3 and R4, with multiple scales caused by pod growth variations and different imaging angles. To resolve this issue, we added an FPN module, which extracts multi-scale features from images, enhances semantic information, and strengthens the detection of small target, multi-scale objects.

Suitable anchor boxes can effectively improve accuracy, for example, Mosley et al. (2020) improved detection accuracy by 7% by adjusting anchor boxes based on annotation data, while Zhang et al. (2020) also achieved significant results by modifying anchor boxes in a tomato-disease-detection algorithm based on object detection. This indicates that adjusting the anchor box size can effectively improve the model detection effectiveness. Consequently, we used *k*-means to cluster the normalized labeled bounding boxes in the training set by adjusting the number and size of anchor boxes to make them better fit the pod shape. *k*-means is a common clustering algorithm that allows the input samples to be clustered by similar characteristics into one class. The distance of all points to all the centroids is calculated, and the closest centroid is assigned to the cluster it represents. After one iteration, the centroids are recalculated for each cluster class, and the closest centroid for each point is found again. The procedure is repeated until there is no change in the cluster class for two iterations before and after. Generally, the distance formula selects the Euclidean distance, whereas, here, we take formula (1) to calculate the distance as follows.

$$d(\text{box, centroid}) = 1 - \text{IOU}(\text{box, centroid}) \quad (1)$$

In Eq. 1, intersection over union (IOU), a concept often used in object detection, is used, where it refers to the ratio of the intersection and merge between the box and the centroid. A complete overlap, i.e., a ratio of 1, is ideal, where box denotes the sample box and centroid is the chosen centroid. The larger the overlap area between the sample box and the chosen centroid, the larger its IOU, and so the smaller the $1 - \text{IOU}$, the smaller the distance of the sample from the centroid. To perform clustering, the distance was calculated according to this formula.

Figure 2C shows the clustering plots under different *K* values derived by *k*-means, where the *K* value refers to the number of anchor boxes, and line plots of the relationship between the number of different anchor boxes and Mean IOU are obtained, as in **Figure 2D**. By investigating the anchor boxes, we found that the Mean IOU changes slowly at 14–18 and almost stops increasing. Therefore, the chosen number of anchor boxes was

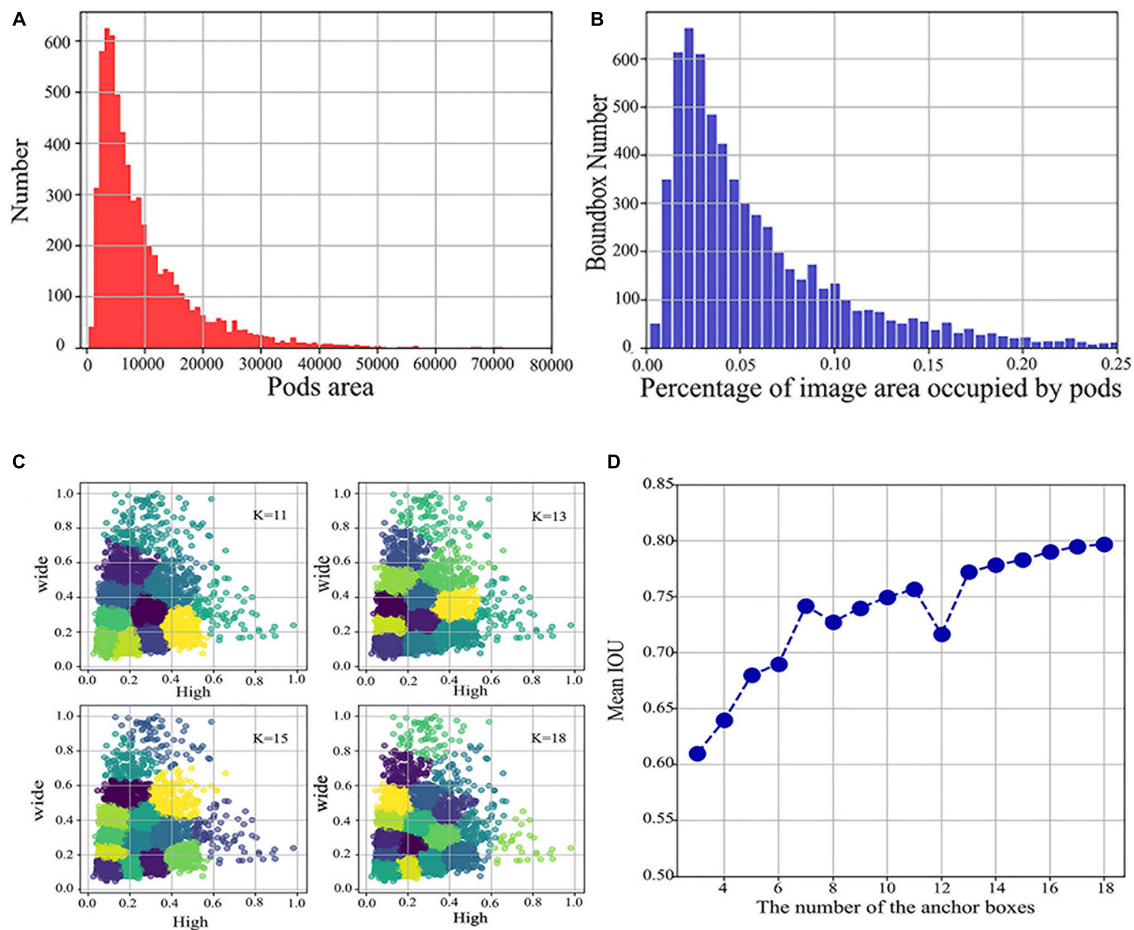


FIGURE 2 | Distribution of pods area and percentage and determining the number of anchor boxes versus the aspect ratio. **(A)** Distribution of the number of different pods areas. **(B)** Distribution of the percentage of the image area occupied by different pods. **(C)** Plots of clustering effects for different K values. **(D)** A plot of the number of different anchor boxes versus Mean IOU.

18, and all the obtained similar-sized aspect ratios were summed to calculate the mean, resulting in three sets of aspect ratios for testing of 0.75, 1.8, and 3.2.

With the above adjustments, we can obtain our fusion model. This is realized by connecting the flower recognition and counting model, and the pod recognition and counting model in series. In the process of image data acquisition, a single plant is taken as the object, so all nodes of the same plant are identified. After identification, a data summary is made. The input image first goes through the flower recognition and counting model to detect the number of flowers, and then into the pod recognition and counting model to detect the number of pods before finally outputting the number of flowers and pods at each node and generating a CSV file, which is summed to the number of flowers and pods per plant. The entire process is shown in **Figure 3**.

Flower Drop Rate, Pod Drop Rate, and Pod Formation Rate

The formulae for calculating the flower drop rate (FDR), PDR, and pod formation rate (PFR) are the key to the study of soybean

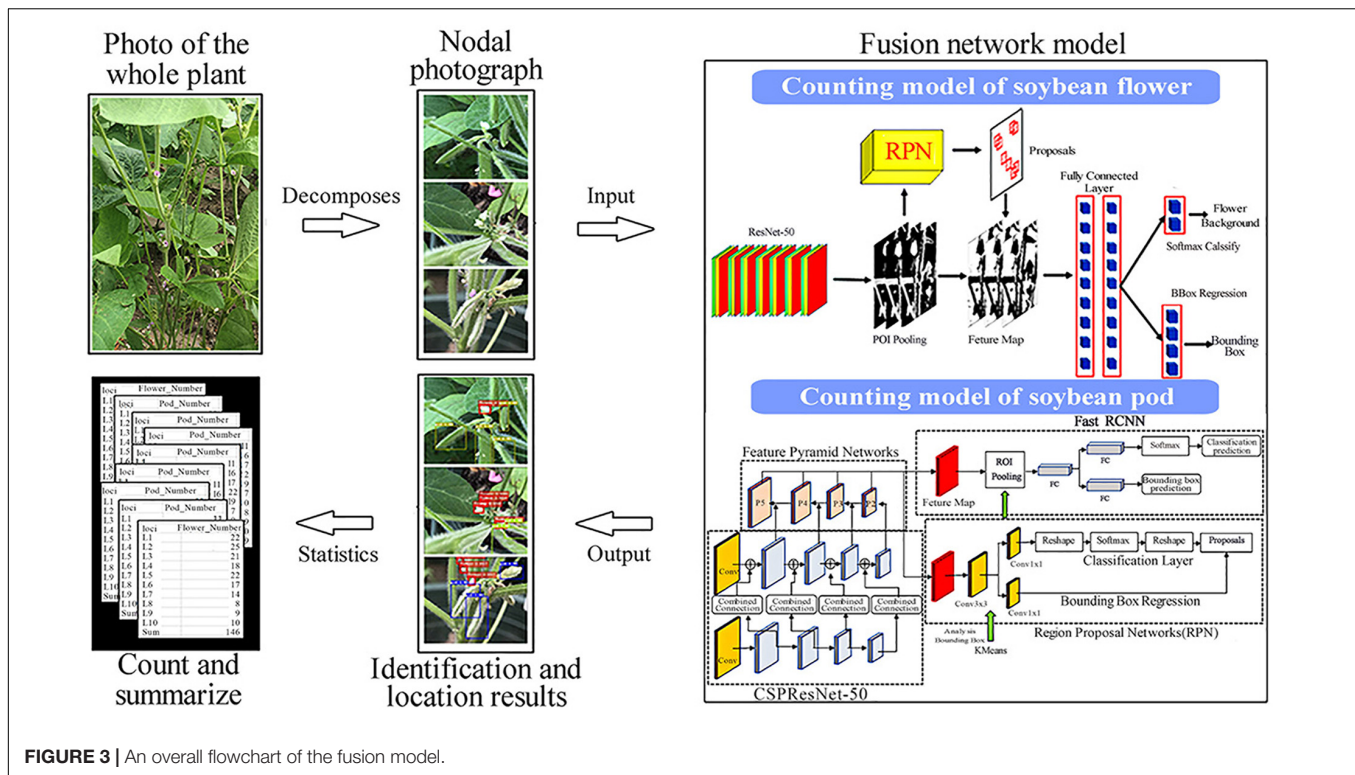
flower and pod drop patterns, and are usually defined as follows, according to the relevant literature (Wang et al., 2014; Xu, 2015).

$$\text{FDR}(\%) = \frac{\text{TNOFD}}{\text{TNOFD} + \text{TNOPD} + \text{TNOFF}} \times 100\% \quad (2)$$

$$\text{PDR}(\%) = \frac{\text{TNOPD}}{\text{TNOFD} + \text{TNOPD} + \text{TNOFF}} \times 100\% \quad (3)$$

$$\text{PFR}(\%) = \frac{\text{TNOFF}}{\text{TNOFD} + \text{TNOPD} + \text{TNOFF}} \times 100\% \quad (4)$$

where FDR = flower drop rate, PDR = pod drop rate, PFR = pod formation rate, TNOPD = total number of flowers dropped, TNOPD = total number of pods dropped, and TNOFF = total number of pods formed. The next step is to determine the soybean plant flower and pod drop rates and explore the soybean flower and pod drop patterns based on the above formulae.



Hardware and Software

To train CNN on the soybean flower image dataset, we used a personal desktop computer with an Intel (R) Core (TM) i9-9900k CPU, NVIDIA Titan XP (12G) GPU, and 64G RAM under the Pytorch and MXNET deep learning frameworks for the Windows operating systems using Python language to train seven networks. While the soybean pod dataset training was based on PaddlePaddle framework programming, the experiments were trained on an NVIDIA GeForce RTX 2080 Ti GPU with 11 Gb RAM, operating system Ubuntu 14.08, CUDA version v10.1, and Cudnn version 7.6.5.

Evaluation Indicators

We used the precision, recall, AP, and mAP values to evaluate the results and performance of the different networks used on the dataset, where the precision, recall, AP, and mAP are determined by the following equations:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (5)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

$$\text{AP} = \sum_{k=1}^N \text{Precision}(k) \Delta \text{Recall}(k) \quad (7)$$

$$\text{mAP} = \frac{\sum_{m=1}^M \text{AP}(m)}{M} \quad (8)$$

where TP (True Positive) indicates that the model correctly identified objects from the defined object region, FP (False Positive) indicates that the background was wrongly identified as an object, and FN (False Negative) indicates that the object was wrongly identified as the background. N is the total image number in the test dataset, M is the total number of categories, m is the number of categories, Precision (k) is the precision value of the k th image, and $\Delta \text{Recall}(k)$ is the change value of the k th image and the $k-1$ th image.

Mean intersection over union was used to evaluate the semantic segmentation model. MIOU calculates the ratio of the intersection and the union of the two sets of true and predicted values and can be calculated as follows.

$$\text{MIOU} = \frac{1}{K+1} \sum_{i=0}^K \frac{TP}{FN + FP + TP} \quad (9)$$

where TP signifies the prediction is correct, the prediction is positive, and the true is positive; FP denotes that the prediction is wrong, the prediction is positive, and the true is negative; FN denotes that the prediction is wrong, the prediction is negative, and the true is positive; K is the total number of categories, and i is the number of categories.

RESULTS

Selection of a Suitable Counting Model

Firstly, we took the semantic segmentation algorithms U-Net, ICNET, Fast-SCNN, DeepLabV3+, and object detection

algorithms Faster R-CNN, YOLOV3, SSD, EfficientDet, and YOLOV5 to detect soybean flowers and pods together, attempting to simultaneously count flowers and pods. For the model evaluation metrics, the object detection algorithm chose the commonly used mAP threshold value of 0.5, and the semantic segmentation algorithm chose MIoU as the evaluation metric. The experimental results are presented in **Table 3**.

The results indicate that the object detection algorithm has a higher detection accuracy than the semantic segmentation algorithm. The highest semantic segmentation algorithm accuracy was 64.42%, and the lowest was 54.01%, both below 70%. The object detection algorithm Faster R-CNN reached an average detection accuracy of 81.7% in the flower and pod counting study, with 85.5% accuracy for flower detection and 77.9% accuracy for pod detection. As can be seen, although the Faster R-CNN has a higher accuracy rate than other object detection algorithms, it still does not meet our requirements for accurately counting flowers and pods simultaneously. Subsequently, the Faster R-CNN algorithm was explored to find a suitable new method, and the Faster R-CNN algorithm was fine-tuned to achieve satisfactory simultaneous counting of flowers and pods. The Faster R-CNN algorithm was then improved for flowers and pods, respectively.

Training and Evaluation of Six Soybean Flowers and Pods Recognition Models

For the soybean flower counting model, we evaluated the performance of the improved Faster R-CNN (ResNet-50) model and five alternatives [Faster R-CNN (VGG16), SSD, EfficientDet, YOLOV3, and YOLOV5]. The learning rate of these six networks starts at 0.0001, with Faster R-CNN (VGG16), SSD, and Faster R-CNN (ResNet-50), making 200 iterations, EfficientDet, YOLOV3, and YOLOV5 use 50 iterations, with the loss values slowly converged to a value near the exact one. We compared the proposed model with five other methods, based on deep learning, under the same experimental configuration, trained with the same training set, and assessed on the test set. The specific experimental evaluation results are presented in **Table 4**.

It can be observed that the highest of the five alternative models is EfficientDet, with an accuracy of 92.83%, but Faster R-CNN (ResNet-50) also outperformed that by 1.53% on the test set. In terms of training time, the Faster R-CNN (ResNet-50) took

longer to train, while the YOLOV3 network took the least time to train. Considering the test accuracy and training time, this study favors Faster R-CNN (ResNet-50) due to its higher detection accuracy and sacrifices training time for accuracy improvement.

For the soybean pod counting model, we evaluated the performance of the improved Faster R-CNN (CSPResNet-50) model and five alternatives [Faster R-CNN (VGG16), SSD, EfficientDet, YOLOV3, and YOLOV5]. The learning rate of these six networks starts at 0.0001, with Faster R-CNN (VGG16), SSD, and Faster R-CNN (CSPResNet-50), making 200 iterations, and EfficientDet, YOLOV3, and YOLOV5 using 50 iterations, with the loss values slowly converged to a value near the exact one. We compared the proposed model with five other deep learning-based methods under the same experimental configuration, trained with the same training set, and assessed on the test set. The specific experimental evaluation results are presented in **Table 4**.

It can be observed that Faster R-CNN (VGG16) is the highest of the five models, with an accuracy of 87.71%, but Faster R-CNN (CSPResNet-50) also outperformed this by 3.29% on the test set. In terms of training time, the Faster R-CNN (CSPResNet-50) took longer to train, while the YOLOV3 and YOLOV5 networks took the least time to train. When considering the test accuracy and training time, this study favors Faster R-CNN (CSPResNet-50), which has a higher detection accuracy and appropriately sacrifices training time for accuracy improvement.

Adjusting the Number and Aspect Ratio of Anchor Boxes

All other things being equal, we used CSPResNet-50 as the Faster R-CNN skeleton network for our experiments and used the *k*-means algorithm to adjust the anchor box aspect ratio in the RPN module to 0.75, 1.8, and 3.2. The effect of aspect ratios on the model is discussed, and the experimental results are shown in **Table 5**. The table shows that following adjustment of the aspect ratio, the overall model detection improved by 0.1% and the small target detection capability improved by 0.5%.

Evaluation of the Soybean Flowers and Pods Counting Results

Approximately, 379 images of soybean flowers were selected along with 539 images of soybean pods, independent of the training sample. The data obtained manually and from Faster R-CNN (ResNet-50) and Faster R-CNN (CSPResNet-50) recognition and counting models were subjected to correlation and error value analyses to evaluate the reliability and accuracy of the models. On this basis, scatter plots and histograms are drawn, as shown in **Figure 4**. The size of the circle in the diagram is related to the number of times a point appears; the more times it appears, the larger the circle. It can be seen from **Figure 4A** that 62.7% of the sample set was identified by the flower model with zero errors, 92.7% with no more than one error, and 99.05% with no more than two errors. The percentage of prediction errors of 0 versus 1 for the pod model in **Figure 4C** is 96%, indicating that the models are still fairly stable. The prediction error value referred to here is the number of differences between the model

TABLE 3 | Experimental results of different deep learning algorithms.

Deep learning algorithms	Models	Evaluation indicators	Accuracy
Object detection	Faster R-CNN	mAP	82.63%
	YOLOV3	mAP	68.9%
	YOLOV5	mAP	76.55%
	SSD	MIoU	54.01%
	EffientDet	mAP	80.84%
Semantic segmentation	Fast-SCNN	mAP	54.01%
	ICNET	MIoU	60.02%
	DeepLabV3+	MIoU	56.32%
	U-Net	MIoU	64.42%

TABLE 4 | Comparison of different object detection algorithms of flowers and pods.

Flower models	Flower training time (h)	Flower detection accuracy (%)	Pod models	Pod training time (h)	Pod detection accuracy (%)
Faster R-CNN (Resnet-50)	7.89	94.36	Faster R-CNN (CSPResNet-50)	7.19	91%
Faster R-CNN (VGG16)	7.97	82.25	Faster R-CNN (VGG16)	6.88	87.71%
SSD	3.02	90.82	SSD	4.03	84.89%
YOLOV3	0.83	63.96	YOLOV3	1.07	50.45%
YOLOV5	0.95	83.40	YOLOV5	1.07	74.25%
EffientDet	0.93	92.83	EffientDet	0.87	75.09%

and manual counts. The coefficient of determination R^2 values for soybean pods, obtained in **Figures 4B,D**, is 0.809 and 0.9046, respectively, indicating that both Faster R-CNN (ResNet-50) and Faster R-CNN (CSPResNet-50) have good soybean flower and pod counting ability and could replace manual counting.

Evaluation of the Fusion Model Count Results

The above exploration reveals that the recognition accuracy of soybean flowers and pods using Faster R-CNN is lower than that of the improved Faster R-CNN (ResNet-50) and Faster R-CNN (CSPResNet-50) separately; thus, a fusion model for soybean flowers and pods is proposed. The detection effect, speed, and accuracy of the fusion model were evaluated. Two soybean plants of two cultivars, DN252 and HN51, were selected as test samples, and the image was fed into a fusion model for flower and pod recognition and counting. The number of flowers and pods per plant was outputted and compared with the manual count, and, on this basis, a linear fit of the fusion model detected pod number versus the actual manual count was plotted, as shown in **Figure 5**. **Figures 5A,B** displays the coefficient of determination R^2 values of 0.9652 and 0.9917 for HN51 and DN252 for flowers, respectively. The coefficient of determination R^2 values of 0.9924 and 0.9872 for pods is shown in **Figures 5C,D**. Both are close to 1, indicating that the fusion model performs well at simultaneously counting soybean flowers and pods.

Study of Flower and Pod Drop Patterns

Analysis of Flower and Pod Variation Patterns During the Reproductive Period

Using the fusion model, we first analyzed the change patterns of flower and pod drops during the soybean reproductive period, mainly in terms of the number of flowering plants, the number of dropped flowers, and the flowering and pod drop timing.

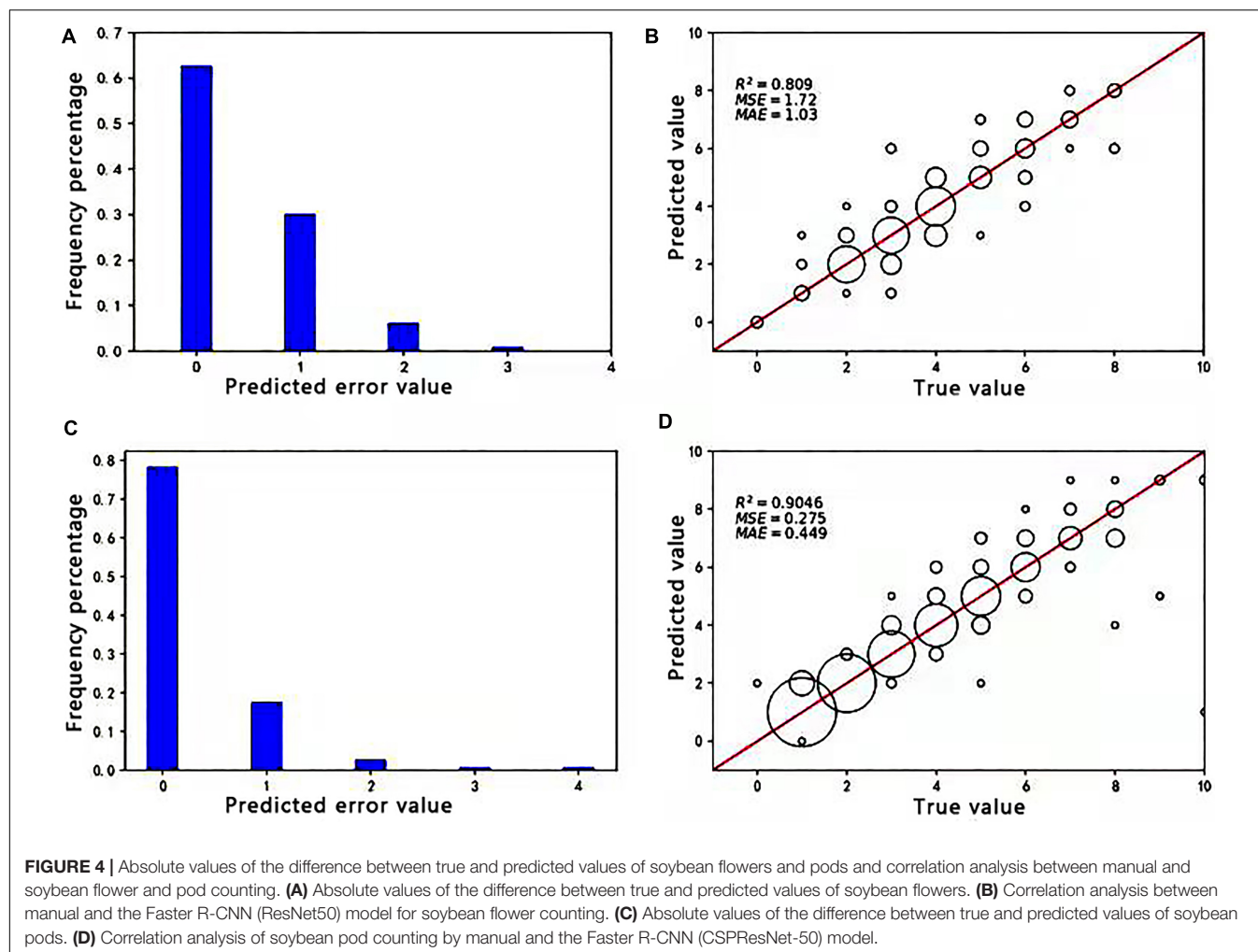
TABLE 5 | Comparison of test results for adjusting the number of anchor boxes and the aspect ratio.

k-means	AP ₅₀	APs	AP _M	AP _L	FPS
No	91%	26.3%	50.9%	66.3%	17
Yes	91.1%	26.8%	52.3%	67.8%	17

The number of DN252 and HN51 flowers and pods was plotted from 27 June 2019 to 30 September 2019, as shown in **Figure 6**. HN51 and DN252 have an average daily flowering count of approximately 7.43 and 4.4 and an average daily flower drop of about 3.975 and 7.86, respectively. It can be seen from **Figure 6** that different soybean cultivars flower and drop at different rates. HN51 blooms slowly, with a high number of flowers; the peak is shifted backward, and the number of flowers reaches its peak after 20 days of flowering and the duration of the fall is about 10–15 days, as shown in **Figures 6A–C**. Meanwhile, DN252 blooms quickly, with a maximum of 24 flowers per day on a single plant; the flower number peak is shifted forward, and the number of flowers is low, but the fall is slower and lasts longer, about 20–26 days, as shown in **Figures 6D,F**. The flowers of both varieties of soybeans last for about a month. The podding timing is similar, with both starting to set around a week after flowering, while HN51 had the fastest podding speed at 3–5 days after peak flowering, gradually reaching a peak, as shown in **Figures 6A–C**. DN252 starts to set pods at the peak flowering time, and pod numbers reach their peak at mid-fall, as shown in **Figures 6D,F**. All pods shed after reaching their peak, and slow flowering varieties are more likely to produce high yields in terms of the final number of pods.

Analysis of the Spatial Distribution Pattern of Soybean Flower and Pod

Next, we investigated the soybean flower and pod spatial distribution pattern. Depending on the podding habit and the number of nodes on the main stem, the distribution can be divided evenly into lower, middle, and upper layers. **Figure 7** shows the total number of flowering plants, pods dropped, and pods formed at each main stem node during the reproductive period. The number of flowers on the branch are calculated at the corresponding node. The flower spatial distribution in both soybean cultivars was found to be mainly in the middle and lower layers, with 40.28% of the flowers on the HN51 branch and 37.05% in the middle layer, while the lower DN252 layer accounts for 38.51%, with 40.62% in the middle layer, as shown in **Figures 7A,B**. The pod drop was observed at all nodes; the HN51 pods drop mainly in the lower and middle zones, while DN252 drops mainly in the middle zone. Both cultivars drop least in the upper zone, as shown in **Figures 7C,D**. Regarding pod formation, the main DN252 pod-forming areas are the lower and



middle layers, while the bulk of HN51 pods forms in the lower layers. This leads to the conclusion that higher pod-forming areas drop more pods and lower pod-forming areas drop fewer pods, as shown in **Figures 7E,F**.

Analysis of Flower and Pod Drop Patterns in Soybeans

The soybean reproductive cycle can be divided into two major growth periods: the first being the V period (Vegetative) and the second being the R period (Reproductive), which can, in turn, be divided into eight reproductive periods: the start of flowering (R1), full bloom (R2), podding starting (R3), full podding (R4), beginning to seed (R5), full seed (R6), beginning to mature (R7), and fully mature (R8; Fehr et al., 1971). The time period covered by this study started at R1 and continued until R8, whereas in previous flower and pod drop studies, rough statistics were presented from the soybean reproductive period, without dividing it into specific reproductive stages. Since it is more convenient to investigate flower and pod drop phenotypes with the proposed fusion model, this is used to conduct statistical analysis during specific reproductive stages. Firstly, the flower drop percentage data between reproductive zones are shown

in **Table 6**, where “period” refers to each soybean reproductive period, “sample” refers to the observed soybean material, and “percentage” denotes the number of flowers dropped by the sample in a given period, as a percentage of the number of flowers dropped during the whole reproductive period. It is found that the FDR varies between cultivars, and the flower drop percentage varies between reproductive stages, varying from time to time but gradually increasing overall. The lowest drop percentage rate is between R1 and R2, with a mean of only 1.9%, while the flower drop peaks between R5 and R6, where it reaches as high as 40%. As the pods grow and develop, the flower drop gradually increases to an average drop rate of 62.49%.

Table 7 displays the PDR statistics in different reproductive zones, where “period” refers to each soybean reproductive period, “sample” refers to the observed soybean sample, and “percentage” is the number of flowers dropped by a soybean sample in a given period as a percentage of the total number of pods dropped within the entire reproductive period. The data indicate that the PDR varies in different reproductive zones, the pod drop in each reproductive zone, with a trend of increasing and then decreasing with soybean growth. During pod growth, development, and seed filling, the pod drop proportion increases significantly, and the

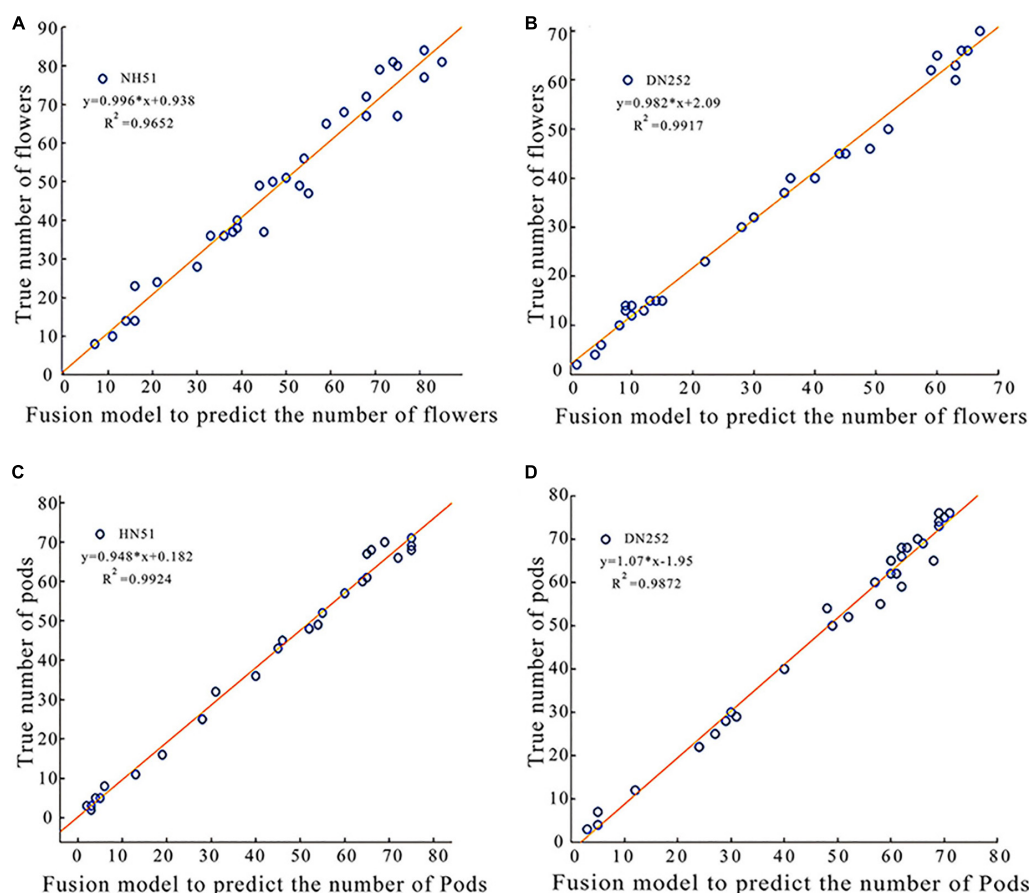


FIGURE 5 | Plots of linear correlation between the fusion model predicted soybean flowers and pods and true values of manual counting. **(A)** A plot of linear correlation between the fusion model-predicted soybean flowers of variety HN51 and true values of manual counting. **(B)** A linear correlation plot of the fusion model predicted the true value of soybean flowers and manual counting for variety DN252. **(C)** A plot of linear correlation between the fusion model-predicted soybean pod of variety HN51 and the true value of manual counting. **(D)** A plot of linear correlation between the fusion model-predicted soybean pod of variety DN252 and the true value of manual counting.

pod drop peaks at the same time as the flower drop peaks, both within R5–R6. Consequently, pod development not only leads to flower drop but also affects the pods themselves. The overall PDR averaged 54.87%.

DISCUSSION

Recognition of Soybean Flowers and Pods in Certain Scenarios

Wan and Goudos (2020) utilized Faster R-CNN in the field of fruit image recognition, and the results indicate that it had good, robust detection results in various complex scenarios. Quan et al. (2019) used the Faster R-CNN model to detect maize seedlings in natural environments and demonstrated high performance under various conditions, including full cycle, diverse weather, and multiple angles, which again proved that Faster R-CNN possesses good robustness in complex environments. In this paper, Faster R-CNN (ResNet-50) and Faster R-CNN (CSPResNet-50) models

were used to detect flowers and pods in different scenarios, and the detection results are presented in **Figure 8**.

The five main confounding factors in soybean flower detection are different flower colors, too small flowers, flower clusters, and overlapping flowers. In this study, the soybean flowers were purple and white, and the different colors interfered with the model. Raindrops on the stems and leaves are extremely similar in texture to white flowers when exposed to sunlight, which can cause the model to misinterpret images. In terms of soybean flower size, most are very small targets that are difficult to identify, leading to model miscues. Furthermore, the flowers are mostly clustered together and are indistinguishable, so some may be missed by the model. When the flowers overlap, those at the back are overshadowed by those at the front, and the model may only detect the flowers at the front and miss those at the back. As can be observed in **Figure 8**, the Faster R-CNN (ResNet-50) model has good flower detection and robustness in all experimental scenarios, as shown in **Figures 8A–E**.

In addition, as soybean pods are a very similar color to the leaves and stems, the pods may be identified as leaves, etc., leading

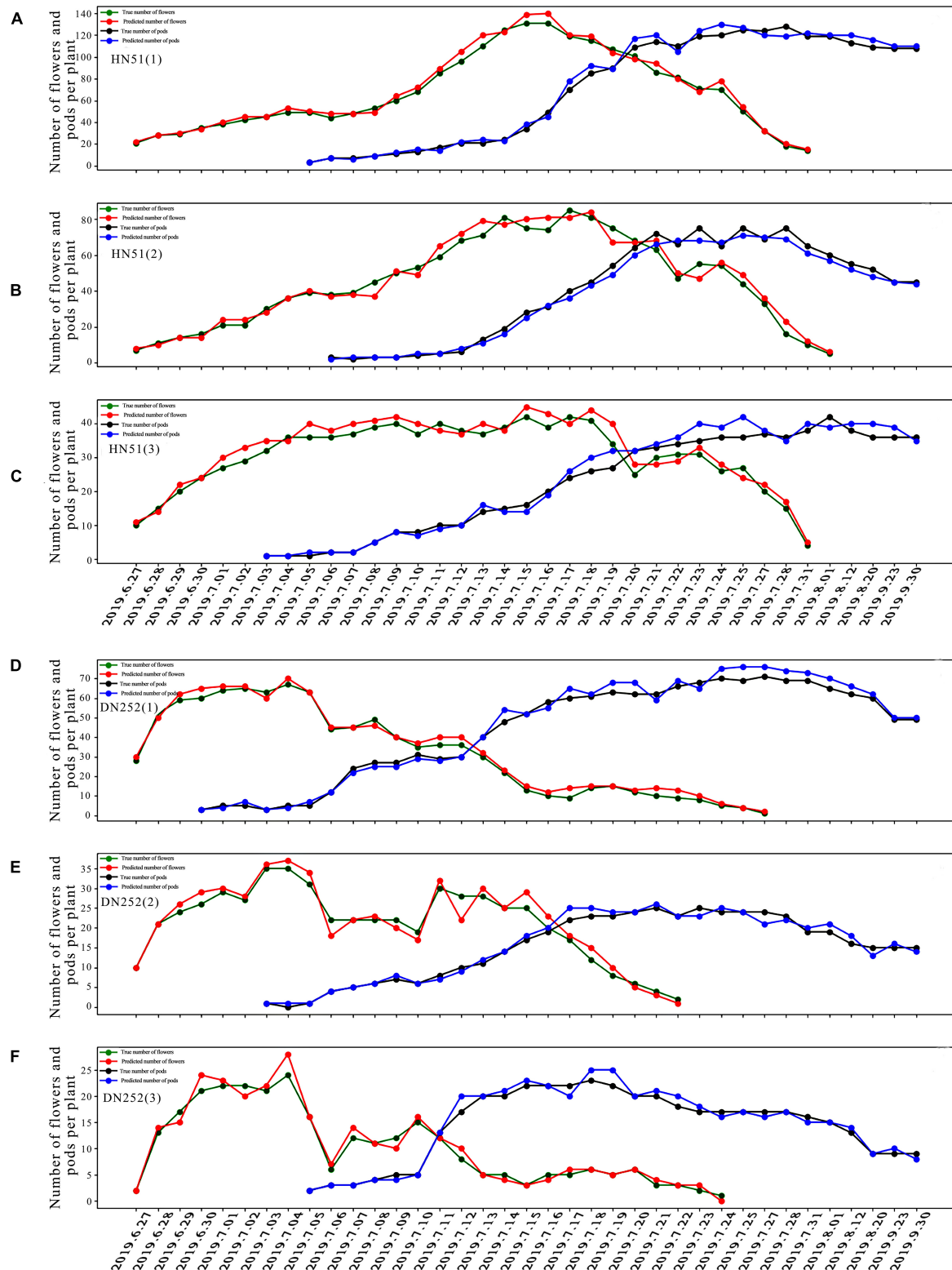


FIGURE 6 | Plots of flower and pod counting over time for three plants of soybean cultivars HN51 and DN252. (A–C) Plots of the number of flowers and pods of three plants of variety HN51 over time. (D–F) Plots of the number of flowers and pods of three plants of variety DN252 over time.

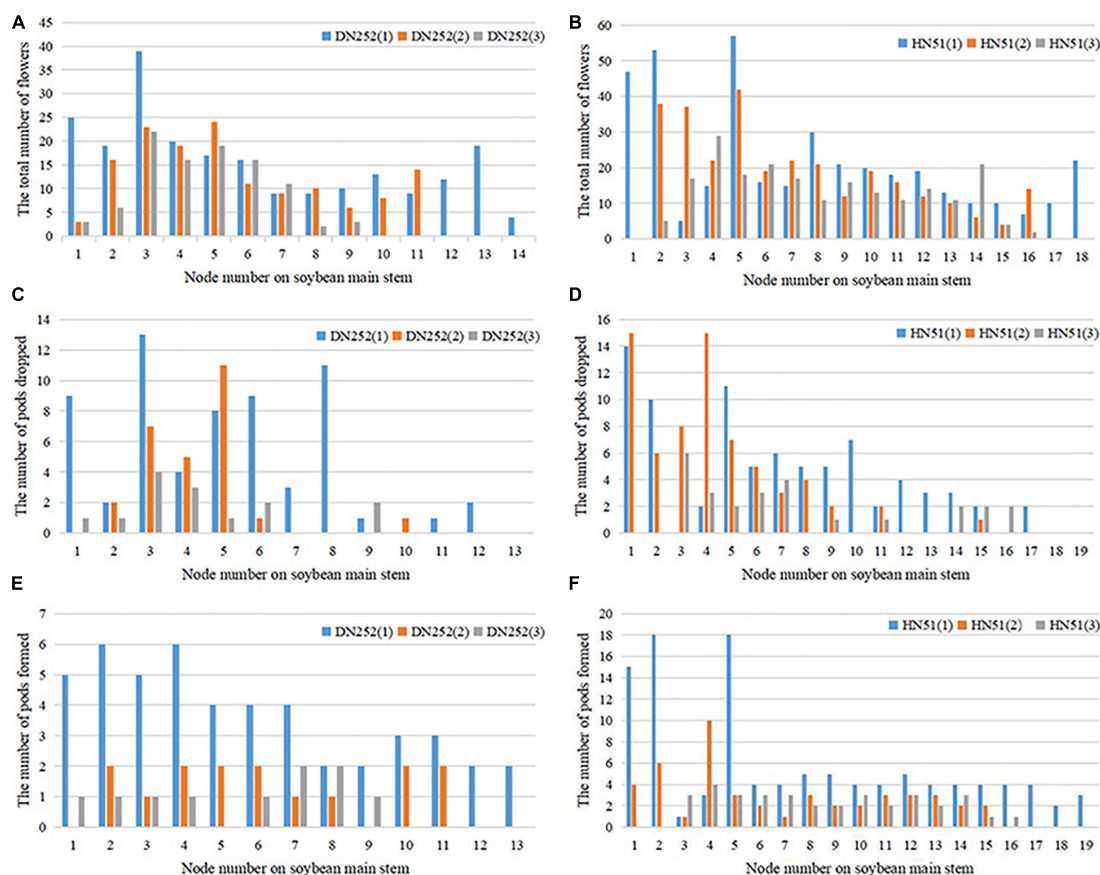


FIGURE 7 | Distribution of total flowers number, number of pods dropped, and number of pods formed at the main stem nodes of the two cultivars DN252 and HN51. **(A)** Distribution of the total number of flowers at the main stem nodes of the DN252 variety. **(B)** Distribution of the total number of flowers at the main stem nodes of variety HN51. **(C)** Distribution of the number of pods dropped at the main stem node of variety DN252. **(D)** Distribution of the number of pods dropped at the main stem node of variety HN51. **(E)** Distribution of the number of pods formed at the main stem node of variety DN252. **(F)** Distribution of the number of pods formed at the main stem nodes of variety HN51.

TABLE 6 | The flower drop rate at different reproductive stages.

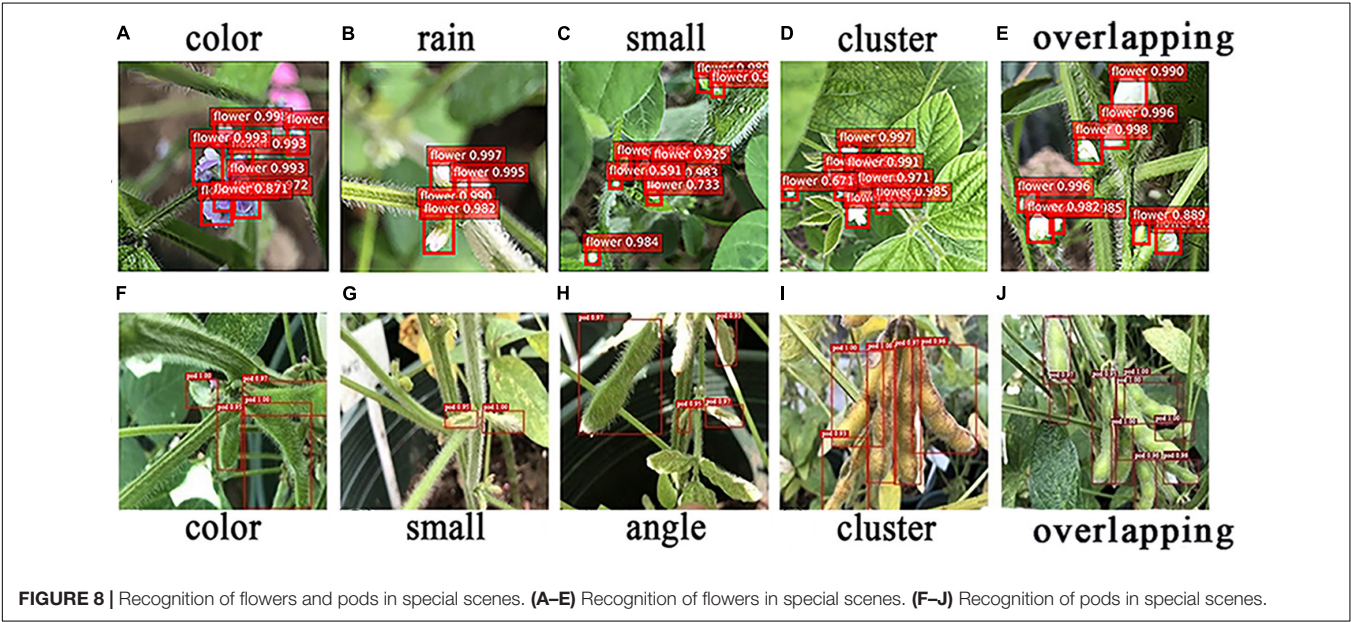
Period								
Percentage								
Sample	R1~R2	R2~R3	R3~R4	R4~R5	R5~R6	Number of flowers dropped	Total number of flowers	Flower drop rate
DN252(1)	3.62%	12.24%	21.26%	24.88%	38%	110	221	49.77%
DN252(2)	3.50%	15.38%	20.29%	16.78%	44.05%	101	143	70.63%
DN252(3)	2.04%	5.11%	36.73%	26.53%	29.59%	74	98	75.51%
HN51(1)	0.26%	2.06%	45.62%	15.72%	36.34%	196	388	50.52%
HN51(2)	1.02%	5.10%	13.95%	29.59%	50.34%	179	294	60.88%
HN51(3)	0.94%	4.25%	28.80%	26.01%	40.00%	142	210	67.62%
Average value	1.90%	7.36%	27.78%	23.25%	40%	133.67	225.67	62.49%

to errors. The different soybean pod periods can lead to the presence of small targets that the model tends to miss. Differences in camera angles can also lead to incorrect judgments, since the soybean pods may have some shape variation due to the angle of the shot, and some pods will also be obscured and remain

unidentified by the model. The soybean pod growth habits, which also bunch together in clusters, can lead to close contact between the pods, and the model may mistake two pods for one, leading to misrecognition. Soybean pods overlap one another, with only a small area being exposed, making it very easy for the model

TABLE 7 | The pod drop rate at different reproductive stages.

Period								
Percentage								
Sample	R3~R4	R4~R5	R5~R6	R6~R7	R7~R8	Number of pods drop	Total number of pods	Pod drop rate
HN51(1)	11.71%	17.11%	23.42%	4.50%	0.00%	111	81	42.18%
HN51(2)	0.00%	11.90%	47.61%	2.38%	2.40%	47	68	59.13%
HN51(3)	8.88%	8.88%	29.17%	12.50%	0.00%	35	33	48.53%
DN252(1)	10.42%	9.90%	18.29%	2.08%	1.56%	48	63	56.75%
DN252(2)	6.96%	13.04%	30.43%	5.22%	3.47%	15	27	64.28%
DN252(3)	7.35%	13.23%	22.06%	1.47%	4.41%	10	14	58.33%
Average value	7.55%	12.34%	28.50%	4.69%	1.97%	44	48	54.87%



to misidentify them. However, the Faster R-CNN (CSPResNet-50) still had good robustness and good detection results for pods under different environmental situations, such as similar color, small targets, shooting angles, clusters, and overlapping interference, as shown in **Figures 8F–J**.

Variation in the Spatial Distribution of Soybean Flowers and Pods Throughout the Reproductive Period

We analyzed the variation in soybean flower and pod spatial distribution throughout the reproductive period, and the results are displayed in **Supplementary Figure 3**. As seen from section “Analysis of the Spatial Distribution Pattern of Soybean Flower and Pod,” the distribution was divided equally into lower, middle, and upper levels based on the pod setting habit and the number of nodes on the main stem. Finally, we studied the pod drop in all layers and periods, with more pods falling during the period R5–R6 and more dropping in the middle and lower layers compared to the upper layers. The pods formed mainly in the lower layers

and dropped more in areas with more pods and less in areas with fewer pods. Soybean flowers are more distributed in the lower and middle layers and fall more where there are more flowers.

We tracked Node 3 and the number of flowers changed from three on 27th June 2019 to four on 3rd July 2019 and then back to three on 12th July 2019. However, one less soybean flower turned into a pod during the period 3rd July 2019 to 12th July 2019. By 17th July 2019, the flowering, flower dropping, and pod formation processes led to three flowers remaining and the pod number increasing to four. One flower was shed by 23rd July 2019, the number of pods remained constant, and one pod was shed between 12th and 20th August 2019. Finally, up until 20th August 2019, the number of pods remained the same before eventually reducing by one.

Pod Formation Traceability Issues

From the above study, we see that the fusion model has a strong ability to identify and count flowers and pods. Subsequently, we attempted to trace the origin of the pods using the fusion

model. This would make it possible to determine which flower a particular pod originates from and to more closely observe the phenotypic changes before and after the flowers are shed. The results are presented in **Supplementary Figure 4**.

We selected a soybean node in order to detect its temporal image using the fusion model. The sequence indicated by the red arrow is the flower to pod tracking record. The results show that the model does not lose tracking with changes in flower shape or color, nor is it affected by tracking errors due to pod color or growth, different camera angles, or complex backgrounds. This demonstrates it possesses a strong, robust detection capability. However, the failing is that the fusion model only mechanically detects flowers and pods in the image and cannot make causal judgments about them, which still requires human-assisted observation. There is also interference from other flowers and pods in the vicinity, so it is not possible to track and locate individual flowers or pods, which is not suitable for dense environments. We are currently unable to solve this problem, but, in the future, we aim to use a counting method that combines an object detection algorithm with the Deep-Sort algorithm (Wojke et al., 2018) to solve the existing problems in the form of numbered tracking counts of tracked target objects. This would enable flower and pod causal judgments while achieving more detailed origin tracking.

CONCLUSION

To investigate soybean flower and pod patterns, we propose a Faster R-CNN-based fusion model for the simultaneous identification and counting of soybean flowers and pods. The main findings are as follows:

1. The improved Faster R-CNN model achieved 94.36 and 91% mAP for soybean flowers and pods, and the results showed that the method can quickly and accurately detect soybean flowers and pods, with strong robustness.
2. The coefficient of determination R^2 between the fusion model's soybean flower and pod counts, and manual counts reached 0.965 and 0.98, indicating that the fusion model is highly accurate in counting soybean flowers and pods.
3. Using the fusion model, we also found the following patterns: firstly, varieties that flower slowly are more likely to produce higher yields; secondly, areas with more pods are more likely to drop pods than areas with fewer pods; thirdly, different varieties have different flower and PDRs and ratios at different stages of fertility, with the flower drop ratio gradually increasing over time and the PDR initially increasing and then decreasing as the soybean grows. The peak range for both flower and pod drops being R5 to R6.

REFERENCES

- Akshatha, K. R., Karunakar, A. K., Shenoy, S. B., and Pai, A. K. (2022). Human detection in aerial thermal images using faster R-CNN and SSD algorithms. *Electronics* 11:1151. doi: 10.3390/ELECTRONICS11071151
- Bock, C. H., Poole, G. H., Parker, P. E., and Gottwald, T. R. (2010). Plant disease severity estimated visually, by digital photography and image analysis, and by

In the future, we will use a combination of object detection algorithms and Deep-Sort algorithms (Wojke et al., 2018) to solve existing counting method issues. We will also locate a platform that allows the automated, high-throughput, and highly accurate acquisition of soybean flower and pod phenotypes to obtain more samples to further demonstrate the patterns we identified.

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 below: https://pan.baidu.com/s/1j-ZE6BHlpVjGay_JqVmOew?password=ate8.

AUTHOR CONTRIBUTIONS

RZ: conceptualization, methodology, and software. XW: data curation and writing—original draft preparation. ZY: software, visualization, and data curation. YQ: data curation and software. HT and ZZ: visualization. ZH: visualization and software. YL: supervision and investigation. HZ: investigation and validation. DX: supervision. QC: investigation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Natural Science Foundation of Heilongjiang Province of China (LH2021C021).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.922030/full#supplementary-material>

Supplementary Figure 1 | A network structure diagram of the target detection model. (A) A network structure diagram of Faster R-CNN. (B) A network structure diagram of SSD. (C) A network structure diagram of EfficientDet. (D) A network structure diagram of YOLOV3. (E) A network structure diagram of YOLOV5.

Supplementary Figure 2 | A network structure diagram of the semantic segmentation model. (A) A network structure diagram of Fast SCNN. (B) A network structure diagram of U-Net. (C) A network structure diagram of ICNet. (D) A network structure diagram of DeepLabV3+.

Supplementary Figure 3 | Structure of the temporal variation in the number of flowers and pods at different nodes of HN51(3) during the period R1-R8.

Supplementary Figure 4 | Tracing the origin of pods.

hyperspectral imaging. *CRC. Crit. Rev. Plant Sci.* 29, 59–107. doi: 10.1080/07352681003617285

- David, E., Madec, S., Sadeghi-Tehran, P., Aasen, H., Zheng, B., Liu, S., et al. (2020). Global wheat head detection (GWHD) dataset: a large and diverse dataset of high resolution RGB labelled images to develop and benchmark wheat head detection methods. *Plant Phenomics* 2020:3521852. doi: 10.34133/2020/3521852

- Du, J., Li, W., Dong, Q., Wang, P., Su, D., Sun, M., et al. (2019). Genetic analysis and QTL mapping on vertical distribution of pod number in soybean. *Soybean Sci.* 38, 360–370. doi: 10.11861/j.issn.1000-9841.2019.03.0360
- Fehr, W. R., Caviness, C. E., Burmood, D. T., and Pennington, J. S. (1971). Stage of development descriptions for soybeans, *Glycine max* (L.) Merrill 1. *Crop Sci.* 11, 929–931. doi: 10.2135/cropsci1971.0011183x001100060051x
- Fei-Fei, L., Deng, J., and Li, K. (2010). ImageNet: constructing a large-scale image database. *J. Vis.* 9, 1037–1037. doi: 10.1167/9.8.1037
- Firdaus-Nawi, M., Noraini, O., Sabri, M. Y., Siti-Zahrah, A., Zamri-Saad, M., and Latifah, H. (2011). DeepLabv3+ encoder-decoder with atrous separable convolution for semantic image segmentation. *Pertanika J. Trop. Agric. Sci.* 34, 137–143.
- Gao, Y., Huang, X., and Yang, H. (1958). Preliminary study on flowers and pods abscission of soybean. *Plant Physiol. commun.* 5, 9–14.
- Ghosal, S., Zheng, B., Chapman, S. C., Potgieter, A. B., and Guo, W. (2019). A weakly supervised deep learning framework for sorghum head detection and counting. *Plant Phenomics* 1:14. doi: 10.34133/2019/1525874
- Gill, T., Gill, S. K., Saini, D. K., Chopra, Y., de Koff, J. P., and Sandhu, K. S. (2022). A comprehensive review of high throughput phenotyping and machine learning for plant stress phenotyping. *Phenomics* 2:156. doi: 10.1007/s43657-022-00048-z
- Hasan, M. M., Chopin, J. P., Laga, H., and Miklavcic, S. J. (2018). Detection and analysis of wheat spikes using convolutional neural networks. *Plant Methods* 14, 1–13. doi: 10.1186/s13007-018-0366-8
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (Piscataway, NJ: IEEE), 770–778. doi: 10.1109/CVPR.2016.90
- Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539
- Li, C., Zou, J., Jiang, H., Yu, J., Huang, S., Wang, X., et al. (2018). Identification and validation of number of pod – and seed – related traits QTLs in soybean. *Plant Breed.* 137, 730–745. doi: 10.1111/pbr.12635
- Li, Y., Jia, J., Zhang, L., Khattak, A. M., Sun, S., Gao, W., et al. (2019). Soybean seed counting based on pod image using two-column convolution neural network. *IEEE Access* 7, 64177–64185. doi: 10.1109/ACCESS.2019.2916931
- Lin, T. Y., Dollar, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. B. T. - (2017). . C. on C. V. and P. R. (CVPR) (2017). Feature Pyramid Networks for Object Detection. *arXiv [preprint]*. arXiv:1612.03144
- Lin, T. Y., Maire, M., Belongie, S., Hays, J., and Zitnick, C. L. (2014). Microsoft COCO: Common Objects in Context. *arXiv [preprint]*. arXiv:1405.0312
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., et al. (2016). “SSD: single shot multibox detector,” in *Computer Vision-ECCV 2016. Lecture Notes in Computer Science*, eds B. Leibe, J. Matas, N. Sebe, and M. Welling (Cham: Springer), 21–37. doi: 10.1007/978-3-319-46448-0_2
- Liu, Y., Wang, Y., Wang, S., Liang, T. T., Zhao, Q., Tang, Z., et al. (2019). CBNet: a novel composite backbone network architecture for object detection. *arXiv [preprint]*. arXiv:1909.03625
- Lu, H., Cao, Z., Xiao, Y., Zhuang, B., and Shen, C. (2017). TasselNet: counting maize tassels in the wild via local contexts regression network. *Plant Methods* 13, 1–17. doi: 10.1186/s13007-017-0224-0
- Ma, G., Liang, Z., and Jin, J. (1960). Investigation on flowers and pods falling of soybean. *Sci. Agric. Sin.* 5.
- Mosley, L., Pham, H., Bansal, Y., and Hare, E. (2020). Image-based sorghum head counting when you only look once. *arXiv [preprint]*. arXiv: 2009.11929
- Mu, Y., Chen, T. S., Ninomiya, S., and Guo, W. (2020). Intact detection of highly occluded immature tomatoes on plants using deep learning techniques. *Sensors (Switzerland)* 20, 1–16. doi: 10.3390/s20102984
- Navab, N., Hornegger, J., Wells, W. M., and Frangi, A. F. (2015). “Medical image computing and computer-assisted intervention – MICCAI 2015,” in *Proceedings of the 18th International Conference Munich, Germany, October 5-9, 2015 Proceedings, Part III*, Vol. 9351, (Cham: Springer), 12–20. doi: 10.1007/978-3-319-24574-4
- Poudel, R. P. K., Liwicki, S., and Cipolla, R. (2020). Fast-SCNN: fast semantic segmentation network. *arXiv [preprint]*. arXiv:1902.04502
- Quan, L., Feng, H., Lv, Y., Wang, Q., Zhang, C., Liu, J., et al. (2019). Maize seedling detection under different growth stages and complex field environments based on an improved faster R-CNN. *Biosyst. Eng.* 184, 1–23. doi: 10.1016/j.biosystemseng.2019.05.002
- Rahmehoonfar, M., and Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. *Sensors (Switzerland)* 17, 1–12. doi: 10.3390/s17040905
- Rampersad, H. (2020). “Developing,” in *Total Performance Scorecard* (London: Routledge), 159–183. doi: 10.4324/9780080519340-12
- Redmon, J., and Farhadi, A. (2018). YOLOv3: an incremental improvement. *arXiv [preprint]*. arXiv:1804.02767
- Riera, L. G., Carroll, M. E., Zhang, Z., Shook, J. M., Ghosal, S., Gao, T., et al. (2020). Deep multi-view image fusion for soybean yield estimation in breeding applications deep multi-view image fusion for soybean yield estimation in breeding applications. *arXiv [preprint]*. arXiv: 2011.07118 doi: 10.34133/2021/9846470
- Sandler, M., Howard, A., Zhu, M., and Zhmoginov, A. (2018). Sandler_mobilenetv2_inverted_residuals_CVPR_2018_paper.pdf. 4510–4520. *arXiv [preprint]*. arXiv:1801.04381v4
- Selvaraj, M. G., Vergara, A., Ruiz, H., Safari, N., Elayabalan, S., Ocimati, W., et al. (2019). AI-powered banana diseases and pest detection. *Plant Methods* 15, 1–11. doi: 10.1186/s13007-019-0475-z
- Singh, A. K., Ganapathysubramanian, B., Sarkar, S., and Singh, A. (2018). Deep learning for plant stress phenotyping: trends and future perspectives. *Trends Plant Sci.* 23, 883–898. doi: 10.1016/j.tplants.2018.07.004
- Singh, A., Jones, S., Ganapathysubramanian, B., Sarkar, S., Mueller, D., Sandhu, K., et al. (2021). Challenges and opportunities in machine-augmented plant stress phenotyping. *Trends Plant Sci.* 26, 53–69. doi: 10.1016/j.tplants.2020.07.010
- Singh, P., Kumar, R., Sabapathy, S. N., and Bawa, A. S. (2010). Functional and edible uses of soy protein products. *Compr. Rev. Food Sci. Food Saf.* 7, 14–28. doi: 10.1111/j.1541-4337.2007.00025.x
- Song, J., Sun, X., Zhang, K., Liu, S., Wang, J., Yang, C., et al. (2020). Identification of QTL and genes for pod number in soybean by linkage analysis and genome-wide association studies. *Mol. Breed.* 40:60. doi: 10.1007/s11032-020-01140-w
- Song, S., and Dong, Z. (2002). Comparative study on blooming sequence and podding habit of soybeans. *Entia. Agric. Sin.* 35, 1420–1423. doi: 10.3321/j.issn:0578-1752.2002.11.022
- Su, L., Dong, Z., and Song, S. (2004). Observation on blooming order of different soybean cultivars. *Rain Fed Crops* 24, 84–85. doi: 10.3969/j.issn:2095-0896.2004.02.009
- Šulc, M., and Matas, J. (2017). Fine-grained recognition of plants from images. *Plant Methods* 13, 1–14. doi: 10.1186/s13007-017-0265-4
- Tan, M., and Le, Q. V. (2019). EfficientNet: rethinking model scaling for convolutional neural networks. *arXiv [preprint]*. arXiv:1911.09070
- Tan, M., Pang, R., and Le, Q. V. (2020). EfficientDet. *arXiv [preprint]*. arXiv:1911.09070
- Thuan, D. (2021). *Evolution of Yolo Algorithm and YoloV5: the State-of-the-Art Object Detection Algorithm*. Bachelor's thesis. Oulu: Oulu University of Applied Sciences, 61.
- Tzatalin, D. (2018). *LabelImg. Git code*. Available online at: <https://github.com/tzatalin/labelImg> (accessed September 22, 2021).
- Ubbens, J., Cieslak, M., Prusinkiewicz, P., and Stavness, I. (2018). The use of plant models in deep learning: an application to leaf counting in rosette plants. *Plant Methods* 14, 1–10. doi: 10.1186/s13007-018-0273-z
- Wan, S., and Goudos, S. (2020). Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput. Netw.* 168:107036. doi: 10.1016/j.comnet.2019.107036
- Wang, B. Q., Dun-Wei, C. I., Zhang, L. F., Wei, L. I., and Ran, X. U. (2010). Research progress of assimilation supply and endogenous hormones signals regulation involved in flower and pod development of soybean. *Soybean Sci.* 29, 878–882.
- Wang, G., and Liu, Z. (2020). *Android Malware Detection Model Based on lightGBM*. Singapore: Springer. doi: 10.1007/978-981-13-9406-5_29
- Wang, H., Sun, X., Yue, Y., Sun, X., Xu, Y., Wang, Y., et al. (2014). Association mapping of flower and pod abscission with SSR markers in northeast spring sowing soybeans. *Soil Crop* 3.
- Wojke, N., Bewley, A., and Paulus, D. (2018). “Simple online and realtime tracking with a deep association metric,” in *Proceedings of the 2017 IEEE International*

- Conference on Image Processing ICIP*, Beijing, 3645–3649. doi: 10.1109/ICIP.2017.8296962
- Wu, W., Liu, T., Zhou, P., Yang, T., Li, C., Zhong, X., et al. (2019). Image analysis-based recognition and quantification of grain number per panicle in rice. *Plant Methods* 15, 1–14. doi: 10.1186/s13007-019-0510-0
- Xiong, H., Cao, Z., Lu, H., Madec, S., Liu, L., and Shen, C. (2019). TasselNetv2: in-field counting of wheat spikes with context-augmented local regression networks. *Plant Methods* 15:150. doi: 10.1186/s13007-019-0537-2
- Xu, R., Li, C., Paterson, A. H., Jiang, Y., Sun, S., and Robertson, J. S. (2018). Aerial images and convolutional neural network for cotton bloom detection. *Front. Plant Sci.* 8:1–17. doi: 10.3389/fpls.2017.02235
- Xu, Y. (2015). *QTL Mapping of Flowers and Pods Abscission Trait in Soybean*. Harbin: Graduate University of Chinese Academy of Sciences.
- Zhang, D., Cheng, H., Wang, H., Zhang, H., Liu, C., and Yu, D. (2010). Identification of genomic regions determining flower and pod numbers development in soybean (*Glycine max* L.). *J. Genet. Genomics* 37, 545–556. doi: 10.1016/S1673-8527(09)60074-6
- Zhang, J., Zhou, T., and Jia, K. (2012). Formation and space-time distribution of flowers and pods for super-high-yielding soybeans. *Soybean Sci.* 5:12.
- Zhang, Y., Song, C., and Zhang, D. (2020). Deep learning-based object detection improvement for tomato disease. *IEEE Access* 8, 56607–56614. doi: 10.1109/ACCESS.2020.2982456
- Zhao, H., Qi, X., Shen, X., Shi, J., and Jia, J. (2018). “ICNet for real-time semantic segmentation on high-resolution images,” in *Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science*, eds V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (Cham: Springer), 418–434.
- Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2000). “IEEE conference on computer vision and pattern recognition, CVPR 2000,” in *Proceedings of the IEEE Computer Society Conference Computer Vision Pattern Recognition*, Seattle, WA. doi: 10.1080/00207160701303912
- Zhao, S., Tang, X., Zhao, X., Feng, Y., Zhao, C., and Zhang, M. (2013). Observation and research on the temporal and spatial distribution of flowering and flowers dropping of soybean. *Sci. Agric. Sin.* 46:12. doi: 10.3864/j.issn.0578-1752.2013.08.003
- 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 Zhu, Wang, Yan, Qiao, Tian, Hu, Zhang, Li, Zhao, Xin and Chen. 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.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: frontiersin.org/about/contact



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership